University of Alberta

Development and Application of Quantitative Proteome Analysis Techniques

by

Chengjie Ji ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Department of Chemistry

Edmonton, Alberta

Spring 2006

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-13994-3
Our file Notre référence
ISBN: 0-494-13994-3

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

for my parents, my wife and my daughter

Abstract

Quantitative proteomics is set to play a pivotal role in the discovery of diagnostic or prognostic protein markers, for the detection of new therapeutic targets, and as a powerful tool to further our understanding of basic biological processes and mechanisms. The realization of these expectations will rely on the development of highly sensitive, efficient and reliable methods for quantitative proteome analysis. This thesis work is devoted to the development and applications of novel strategies for quantitative proteome analysis based on global stable isotope labeling and microbore liquid chromatography (LC) matrix-assisted laser desorption/ionization (MALDI) quadrupole time-of-flight (QqTOF) mass spectrometry (MS).

Compared to conventional LC electrospray ionization (ESI) based quantitative proteomic strategies, LC-MALDI based strategies allow global quantification and selective identification of differentially expressed proteins between two proteome samples, such as two cell lines. It was demonstrated that differential isotope dimethyl labeling of peptides and two-dimensional LC-MALDI QqTOF MS allowed accurate quantification and identification of differentially expressed proteins. Setting a relative abundance ratio of greater than 2-fold was demonstrated to be a very stringent threshold to quantify and identify differentially expressed proteins between two cultured cell lines. LC-MALDI also allowed the use of microbore instead of microcapillary LC for peptide separation, providing higher sample loading, accurate protein quantification due to the increased signal-to-noise ratio of peptide pairs, and the quantification of protein mixtures with a concentration dynamic range of as high as 1×10^4.

It was demonstrated that a novel isotope labeling technique, 2MEGA, involving N-terminal dimethylation (2ME) after lysine guanidination (GA), offered several desirable features, including: simple experimental procedure, the use of inexpensive and commercially available reagents, and negligible isotope effect on reversed-phase separations. In addition to its applicability to quantitative analysis, LC-MALDI MS combined with 2MEGA was successfully used to identify proteins that included polymorphic variants, phosphopeptides and low abundance proteins in complex samples, such as the whey fraction of bovine milk. It was also demonstrated that the enhanced a_1 or a_1-related peaks in both MALDI and ESI tandem mass spectra of 2MEGA labeled peptides provided additional information that reduced false positive identifications.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Liang Li, most sincerely for granting me the opportunity to work with him, and for his invaluable guidance, inspiration and advice throughout my graduate studies. His careful critique and enthusiastic encouragement to press on towards the goal was truly appreciated.

I also would like to thank sincerely the other members of my supervisory committee: Professor Charles A. Lucy and Professor Derrick L.J. Clive, and the other members of my examining committee: Professor Robert E. Campbell, Professor Manijeh Pasdar, and Professor Robert S. Brown, for their active participation during my oral examination and for their thorough reviews and comments on this thesis.

I gratefully acknowledge the participation of other people with whom I collaborated. I especially thank Dr. Laiji Li and Dr. Manijeh Pasdar, of the Department of Cell Biology at the University of Alberta for collaborative work on the squamous carcinoma cell lines: Dr. Laiji Li for his labor in preparing squamous carcinoma cells and doing immunoblotting and immunofluorescence, and Dr. Manijeh Pasdar for her helpful directions and useful discussions in this exciting area of research (Chapter 4). I also thank Mr. Mulu Gebre for his help with strong cation exchange separations and LC-MALDI experiments (Chapter 4), Ms. Hui Dai for her kind help with the data analysis (Chapters 3 and 4), and Ms. Nan Guo for collecting all of the MALDI MS spectra with the Bruker MALDI TOF and part of the MALDI MS/MS spectra with the Qstar instruments (Chapter 5). I am grateful to have worked with Dr. Nan Zhang, and would like to thank her for the great contribution she made with the reproducibility study (Chapter 6). I thank Mr. Andy Lo for his hard work on desalting using solid phase extraction, strong cation exchange separation, and for his help with the data analysis (Chapter 7). I also thank Dr. Sandra Marcus for growing *E.coli* cells and preparing the membrane fraction (Chapter 7).

Special thanks are due to Dr. Nan Zhang, Dr. Zhengping Wang, Dr. Bernd Keller, Dr. Rui Chen, Dr. Hongying Zhong, Dr. Nan Li, Dr. Alan Doucette, Dr. David Craft, Mr. Christopher McDonald and Ms. Jing Zheng for their practical advice and professional teaching of instrument use and protocols. All the other members of Dr. Li's group are

also gratefully acknowledged for their helpful discussions: in no particular order, Dr. Boyan Zhang, Dr. Ken K.-C. Young, Dr. Truong Ta, Dr. Huaizhi Liu, Dr. Russell Handy, Dr. Ying Zhang, Ms. Lidan Tao, Ms. Xinlei Yu, and Mr. Bryce Young. I especially thank Ms. Tien Quach for her friendship and encouragement.

I also thank Mr. Michael Carpenter for his thorough proof-reading of this thesis. With his help I eliminated many questionable grammatical endeavors into the English language.

I would like to thank the personnel in the general, purchasing, and post offices, and the electronics and machine shops at the Department of Chemistry of the University of Alberta. Without their kind help, this research work would not have been possible.

Special thanks go to Dr. Randy Whittal, the Coordinator of the Mass Spectrometry Facility in the Department of Chemistry, University of Alberta, for his professional training on how to use the QTOF instrument and also for his helpful suggestions.

I would like to thank the Department of Chemistry and the University of Alberta for providing me with the opportunity to learn and grow as a scientist, and for providing a safe and well-equipped facility to conduct my experiments, the Alberta Heritage Foundation for Medical Research for a full-time graduate studentship, and Genome Canada for providing funding for my research projects.

Finally, I especially thank my wife, Yankun for her love, support, patience and encouragement over the course of my graduate studies, my daughter, Kaylee, for the happiness she has brought to me since she was born several months ago, and my parents-in-law for their grateful help in taking care of her.

Table of Contents

Chapter 1 1
Introduction: 1
Modern Mass Spectrometry Combined with Separation Techniques for Qualitative and Quantitative Proteomics 1
1.1 Ionization Methods 1
1.1.1 MALDI 2
1.1.2 Electrospray ionization 5
1.2 Mass Analyzers 8
1.2.1 TOF 8
1.2.2 Quadrupole 10
1.2.3 Ion Trap 14
1.3 Separation Techniques 15
1.4 Protein Identification 18
1.4.1 Peptide Mass Fingerprinting 18
1.4.2 Database Searching with Tandem Mass Spectrometric Data 20
1.5 Post-translational modifications 25
1.6 Quantitative proteomics 26
1.6.1 The Principle 27
1.6.2 Isotopic Labeling 29
1.6.2.1 In Vivo 29
1.6.2.2 In Vitro 32
1.6.2.3 Labeling during proteolysis (^{18}O labeling) 34
1.6.3 *De Novo* Sequencing 36
1.7 Objectives of the thesis 36
1.8 Literature Cited 37

Chapter 2 45
A New Mass Spectrometry Method for the Study of Post-Translational Modifications of Low-Mass Proteins 45

2.1 Introduction 45
2.2 Experimental 47
2.2.1 Chemicals and Reagents 47
2.2.2 Cell Growth and Protein Extraction 47
2.2.3 Fractionation of Proteins from the Soluble *E. coli* Whole Cell Lysates by RP-HPLC 48
2.2.4 In-solution Digestion 49
2.2.5 LC ESI MS/MS 49
2.2.6 1D SDS PAGE 50
2.2.7 In-gel digestion 50
2.2.8 MALDI MS 51
2.2.9 MALDI MS/MS 52
2.2.10 Possible Matches in Public Proteome Database 53
2.2.11 MALDI MS and MALDI MS/MS Mass Spectra Interpretation and Database Searching 53
2.3 Results and Discussion 54
2.4 Conclusions 72
2.5 Literature Cited 72

Chapter 3 75
Quantitative Proteome Analysis Using Differential Stable Isotopic Labeling and Microbore LC-MALDI MS and MS/MS 75
3.1 Introduction 75
3.2 Experimental 80
3.2.1 Chemicals and Reagents 80
3.2.2 Preparation of Control Protein Mixture Digests 80
3.2.3 *E. coli* Protein Extraction and Digestion 81
3.2.4 Preparation of Methyl Esters 82
3.2.5 Reductive Amination 83
3.2.6 Protein Mixtures for Concentration Dynamic Range Studies 83
3.2.7 Microbore LC-MALDI QqTOF Mass Spectrometric Analysis 84

3.2.8 Protein Identification from MS/MS Data 85
3.2.9 LC-ESI MS 85
3.3 Results and Discussion 86
3.3.1 Comparison of Labeling Methods 88
3.3.2 Isotope Effect 96
3.3.3 Quantification of Low Abundance Proteins 99
3.3.4 *E. coli* Whole Cell Extracts 102
3.4 Conclusions 106
3.5 Literature Cited 107

Chapter 4 129

Identification and Quantification of Differentially Expressed Proteins in E-Cadherin Deficient SCC9 Cells and SCC9 Transfectants Expressing E-Cadherin by Dimethyl Isotope Labeling, LC-MALDI MS and MS/MS 129
4.1 Introduction 129
4.2 Experimental 132
4.2.1 Chemicals and Reagents 132
4.2.2 Cell culture and sample preparation 133
4.2.3 Protein extraction 133
4.2.4 Acetone precipitation and in-solution protein digestion 134
4.2.5 Dimethyl labeling of the protein digests 135
4.2.6 Cation exchange chromatography 135
4.2.7 Microbore LC-MALDI QqTOF 136
4.2.8 Relative quantification of peptide pairs 137
4.2.9 Protein identification from MS/MS data 137
4.2.10 Electrophoresis and Western blotting 138
4.2.11 Immunofluorescence 139
4.3 Results and Discussion 139
4.4 Literature Cited 155

Chapter 5 161

Differential Dimethyl Labeling of N-termini of Peptides after Guanidination for Proteome Analysis 161
5.1 Introduction 161
5.2 Experimental 166
5.2.1 Chemicals and Reagents 166
5.2.2 Tryptic Digestion of Myoglobin 166
5.2.3 Preparation and Digestion of Whey Fractions of Skim Milk 167
5.2.4 2MEGA Labeling 168
5.2.5 MALDI MS 169
5.2.6 Microbore LC-MALDI QqTOF Mass Spectrometric Analysis 170
5.2.7 Protein Identification from MS/MS Data 171
5.2.8 LC-ESI MS 172
5.3 Results and Discussion 173
5.3.1 2MEGA Labeling Strategy 173
5.3.2 Isotope Effect 179
5.3.3 CID of the 2MEGA Labeled Peptides 180
5.3.4 Quantification and Identification of Phosphorylated Peptides and Low Abundance Proteins in Milk 182
5.3.5 *De Novo* Peptide Sequencing with 2MEGA Differential Isotope Labeling 190
5.4 Conclusions 195
5.5 Literature Cited 197

Chapter 6 202
Evaluation of Quantitative Reproducibility Using 2MEGA Labeling and LC-MALDI MS202
6.1 Introduction 202
6.2 Experimental 206
6.2.1 Chemicals and Reagents 206
6.2.2 Cell Growth 206
6.2.3 Membrane Protein Isolation and Tryptic Digestion 207
6.2.4 2MEGA Labeling 208

6.2.5 Desalting Using Solid Phase Extraction Cartridge 208
6.2.6 Cation Exchange Chromatography 209
6.2.7 Microbore LC-MALDI QqTOF Mass Spectrometric Analysis 209
6.2.8 Peak Pair Detection and Abundance Ratio Calculation 210
6.2.9 Nano-LC ESI QTOF Mass Spectrometric Analysis 211
6.2.10 Protein Identification from MS/MS Data 212
6.3 Results and Discussion 212
6.4 Conclusions 221
6.5 Literature Cited 222

Chapter 7 232
Large-Scale Evaluation of the Effect of 2MEGA Labeling for Membrane Proteome Analysis Using Nano-LC ESI QTOF 232
7.1 Introduction 232
7.2 Experimental 236
7.2.1 Chemicals and Reagents 236
7.2.2 Cell Culture and Membrane Preparation 237
7.2.3 Protein Digestion 238
7.2.4 2MEGA Labeling 239
7.2.5 Desalting Using Solid Phase Extraction Cartridge 240
7.2.6 Cation exchange chromatography 241
7.2.7 Nano-LC ESI QTOF Mass Spectrometric Analysis 241
7.2.8 Protein Identification from MS/MS Data 242
7.2.9 Hydropathy Calculation 243
7.3 Results and Discussion 243
7.3.1 Fragmentation of Electrosprayed 2MEGA Labeled Peptides 244
7.3.2 The Effect of Instrument Settings on Database Searching 254
7.3.3 The Effect of 2MEGA Labeling on the Proteome Analysis 257
7.3.4 Identification of Membrane Proteins in *E.coli* Membrane Fraction 259
7.4 Conclusions 264
7.5 Literature Cited 266

Chapter 8 290
Conclusions and Future Work 290

List of Figures

Figure 1.1 Schematic diagram of MALDI process. 3

Figure 1.2 Chemical structure of two common MALDI matrices. 3

Figure 1.3 Schematic diagram of ES process..6

Figure 1.4 Schematic diagram of a reflectron MALDI TOF. 10

Figure 1.5 Schematic diagram of a quadrupole mass analyzer. 11

Figure 1.6 Mass filter created by a quadrupole mass analyzer. (A) Negative rods create a low-pass mass filter. (B) Positive rods create a high-pass mass filter. (C) Combining negative and positive sets of rods into a quadrupole arrangement overlaps the two mass filter regions, creating an area of mutual stability which allows ions of a certain m/z to pass...12

Figure 1.7 Schematic diagram of the tandem QTOF mass spectrometer. 13

Figure 1.8 Schematic diagram of an ion trap mass analyzer. 14

Figure 1.9 Peptide fragmentation nomenclature. (A) Nomenclature for peptide fragment ions that form via cleavage of bonds along the peptide backbone. (B) Example structure for b and y ions. (C) Immonium ion. Amino acids generate immonium ions to differing extents and some not at all. 22

Figure 1.10 Schematic representation of different strategies for stable-isotope protein or peptide labeling for quantitative proteomics. Note: for the in vitro labeling strategy, some stable isotope reagents (e.g. ICAT) can label proteins instead of peptides. 30

Figure 2.1 Schematic representation of the strategy used for annotating the origins of apparent masses observed in MS-derived *E. coli* 47076 protein mass database. Proteins identified by tandem MS/MS in each HPLC fraction were assigned to those apparent masses observed in the MALDI MS spectrum of the same fraction. 55

Figure 2.2 Chromatogram of cell extracts of *E.coli*. 57

Figure 2.3 Proteome map of HPLC fraction 36 of *E. coli* cell extracts. 64

Figure 2.4 Fragmentation of OMPA observed in MALDI MS. (A) MALDI MS spectrum of HPLC fraction 36 of *E. coli* cell extracts. (B) Peptide map of tryptic digest of the band indicated in Figure 2.3. Peptides matched with that of OMPA are labeled with an asterisk (*). (C) LC MS/MS spectrum of peptide sequence of NNSLSQEVQNAQHQR. (D) LC MS/MS spectrum of peptide sequence of SLEVFEK. 65

Figure 2.5 Hydroxylation of arginine in 50S ribosomal protein L16 observed in MALDI MS. (A) MALDI MS spectrum of HPLC fraction 50 of *E. coli* cell extracts. (B) LC MS/MS spectrum of peptide with sequence of NNSLSQEVQNAQHQR. (C) LC MS/MS spectrum of peptide with sequence of SLEVFEK. 69

Figure 2.6 Loss of MTM in hypothetical protein YIIU_ECOLI observed in MALDI MS. (A) MALDI MS spectrum of HPLC fraction 49 of *E. coli* cell extracts. (B) MALDI MS/MS spectrum of peptide sequence of VLYEMDGVPEELAR. (C) MALDI MS/MS spectrum of peptide sequence of VFEPDKPITEKPLAVR. 71

Figure 3.1 Workflow for quantitative protein analysis by combining global differential stable isotope labeling with microbore LC-MALDI QqTOF mass spectrometry. 87

Figure 3.2 Quantification and identification of a peptide generated from a mixture of standard proteins. (A) Expanded MALDI mass spectrum. The relative quantities of the protein in the d(0)- and d(3)-labeled mixtures were obtained by comparing the peak intensities of the monoisotopic peaks at m/z values of 1544.642 and 1553.695, respectively. (B) MALDI MS/MS spectrum of the peptide ion at m/z 1544.642. The spectrum obtained was matched to the sequence HPGDFGADAQGAMTK (calculated m/z 1544.773) from MYG_HORSE, where the carboxylic acid groups in D and C-terminus were modified with the d(0)- form of the methanol reagent. 92

Figure 3.3 Quantification and identification of a peptide pair with significant overlap of isotope envelops. The d(0)/d(4) value was determined by comparing the corrected signal intensities of monoisotopic peaks of at m/z of 2175.04 and 2179.08 after the deconvolution and background subtraction. The d(4)-labeled peptide at m/z of 2179.08 was chosen for CID and successfully identified as GWQVPAFTLGGEATDIVVMR. 95

Figure 3.4 Quantification and identification of a low abundance protein. (A) The MALDI mass spectrum obtained for the fraction at 18 min from the microbore column separation. (B) The expended view indicated by rectangular in (A). (C) and (D) MALDI MS/MS spectra for a pair of peptides at m/z 768.461 and 780.553. 101

Figure 3.5 (A) MALDI mass spectrum obtained for the fraction at 45 min from the microbore column separation. (B) and (C) MALDI MS/MS spectra for a pair of peptides at m/z 1344.597 and 1352.667. (D) Overlaid MS/MS spectra of (B) and (C). 103

Figure 4.1 Two-dimensional LC MALDI MS and MS/MS analysis of differentially expressed proteins between SCC9 and SCC9-E cells. (A) SCX chromatogram of d(0)- and d(2)-formaldehyde labeled peptides mixture. UV absorbance was recorded at 214 nm. (B) RP chromatogram of d(0)- and d(2)-formaldehyde labeled peptides in one SCX fraction collected between 19 and 20 min as indicated in (A). UV absorbance was recorded at 214 nm. (C) MALDI mass spectrum obtained for the fraction collected at 28 min. The inset shows the expanded view of the peptide pair as indicated by the rectangular. (D) and (E) MALDI MS/MS spectra for a pair of peptides at m/z 1043.54 and 1051.59. (F) Overlaid MS/MS spectra of (D) and (E)...142

Figure 4.2 Validation of differential expression of selected proteins in E-cadherin-expressing SCC9 cells. (A) Equal amounts of total cellular proteins from SCC9 and SCC9-E cells were resolved on 6% ($p120^{ctn}$, α-cat, caldesmon and tubulin) or 10% (calreticulin, colligin, peroxiredoxin and tubulin) SDS-gels and processed for immunoblotting with various antibodies as described in Materials and Methods. To confirm equal amounts of protein the same blot processed with tubulin antibodies. (B) Gels from A were scanned and quantified using NIH Image analysis software. Histograms were constructed by normalizing the amount of each protein band to the tubulin in the same cell line relative to the SCC9 cells. 151

Figure 4.3 Subcellular distribution of selected differentially expressed proteins in E-cadherin-expressing SCC9 cells. SCC9 and SCC9-E cells were grown on glass coverslips and processed for immunofluorescence microscopy with $p120^{ctn}$, α-catenin, caldesmon, calreticulin, colligin or peroxiredoxin antibodies as described in Materials and Methods. 152

Figure 5.1 Overview of the 2MEGA differential isotope labeling strategy for quantitative proteomic analysis and peptide sequencing. (A) The guanidination reaction selectively modifies the ε-amino group of the lysine residue and the N-terminus of the resultant peptide is isotopically labeled with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde. (B) Workflow for quantitative protein analysis and de novo peptide sequencing by combining the 2MEGA differential isotope labeling with microbore LC-MALDI QqTOF MS. 174

Figure 5.2 MALDI spectra of (A) native, (B) guanidinated, (C) N-terminal d(0),12C-formaldehyde labeled, and (D) N-terminal d(2),13C-formaldehyde labeled myoglobin digest...175

Figure 5.3 Selected MALDI MS/MS spectra of (A) native, (B) N-terminal d(0),^{12}C-formaldehyde labeled, (C) N-terminal d(2),^{13}C-formaldehyde labeled tryptic peptide YKELGFQG derived from myoglobin, and (D) overlaid MS/MS spectra of (B) and (C). 181

Figure 5.4 Accurate quantification and unambiguous identification of a phosphopeptide pair from a protein digest of bovine milk whey fraction. (A) and (B) Partial

MALDI mass spectra obtained from two consecutive LC fractions on the MALDI target. (C) and (D) MALDI MS/MS spectra for a pair of peptides at m/z 2131.80 and 2137.84 (FQsEEQQQTEDELQDK)..........................185

Figure 5.5 (A) Partial MALDI mass spectrum obtained from the fraction at 43 min from the microbore column separation of a protein digest of bovine milk whey fraction. (B) and (C) MALDI MS/MS spectra for a pair of peptides at m/z 1165.58 and 1171.62 along with the summary of MASCOT search results shown in insets. (D) Overlaid MS/MS spectra of (B) and (C). 193

Figure 5.6 An example shows the identification of the new variant in β-lactoglobulin. The peptide sequence was de novo as L(I)I(L)VTQTVK, which corresponds to (17)LIVTQTMK(24) in β-lactoglobulin with amino acid variation occurring at residue 23 (M → V). (A) and (B) MALDI MS/MS spectra for a pair of peptides at m/z 971.59 and 971.60. (C) Overlaid MS/MS spectra of (A) and (B).. `196

Figure 6.1 Workflow for investigation of quantification variations using integrated strategy of N-terminal differential dimethyl labeling and microbore LC-MALDI QqTOF MS. A flowchart for measuring the quantitative reproducibility starting at (A) post proteins level. (B) membrane protein preparation level. (C) cell growth level. 215

Figure 6.2 The measured ratios of all peptide pairs shown in MALDI MS spectra plotted against the protonated peptide masses of d(0), ^{12}C-formaldehyde labeled peptides from quantitative reproducibility analysis starting at post proteins level. (A) and (B) are from paralleled experiments. 218

Figure 6.3 The measured ratios of all peptide pairs shown in MALDI MS spectra plotted against the protonated peptide masses of d(0), ^{12}C-formaldehyde labeled peptides from quantitative reproducibility analysis starting at membrane protein preparation level. (A) and (B) were plotted from duplicated experiments. 219

Figure 6.4 The measured ratios of all peptide pairs shown in MALDI MS spectra plotted against the protonated peptide masses of d(0), ^{12}C-formaldehyde labeled peptides from quantitative reproducibility analysis starting at cell growth level. (A) and (B) were plotted from duplicated experiments.......220

Figure 7.1 Flowchart of the workflow for the investigation of the effect of 2MEGA labeling on the large-scale membrane proteome analysis..........................245

Figure 7.2 ESI MS/MS spectra of (A) 2MEGA labeled and (B) unlabeled GYDDEDILK. * indicates amino group on the side chain of lysine is blocked by guanidination...247

Figure 7.3 ESI MS/MS spectra of various peptides. KAEQWATGLK is shown (A)

List of Tables

Table 2.1 Protein masses of protonated species observed in MALDI MS spectra of HPLC fractions of *E.coli* cell extracts, and tentative protein assignments based on the molecular mass and matched proteins and possible PTMs identified by combining LC MSMS and MALDI MSMS. 58

Table 3.1 Esterification labeling and microbore LC-MALDI QqTOF analysis results of two control protein mixtures....................89

Table 3.2 Dimethylation labeling and microbore LC-MALDI QqTOF analysis results of two control protein mixtures....................90

Table 3.3 Proteins identified and quantified in HPLC fraction 45 of *E. coli* extract samples....................105

Table 3.4 Proteins identified and quantified in *E.coli* extract samples. 112

Table 4.1 List of identified proteins differentially expressed between SCC9 and SCC9-E cells....................145

Table 5.1 Theoretical and experimental MALDI-MS data for myoglobin tryptic digest* 176

Table 5.2 Identification and quantification of proteins in the whey fraction of skim milk. 186

Table 5.3 Theoretical masses of the a_1 ions derived from the twenty amino acid residues without and with N-terminal dimethyl labeling with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde. 192

Table 6.1 Summary of statistical analysis results of quantitative reproducibility at three different levels. 216

Table 6.2 Identification of proteins in membrane protein fraction enriched using MEM-PER kit. 225

Table 7.1 Theoretical masses of the a_1, a_1-NH_3 and a_1-$HN(CH_3)_2$ ions derived from the twenty amino acid residues after 2MEGA labeling. 251

Table 7.2 Comparison of database search results from MASCOT using default ESI-QUAD-QTOF without counting immonium ions and a-series ions and custom modified ESI-QUAD-QTOF counting immonium ions and a-series ions....................255

Table 7.3 Summary of analysis results of identified peptides generated from membrane fraction of *E. coli* cell extract under labeled and unlabeled conditions......258

Table 7.4 Subcellular localizations of proteins identified from unlabeled and 2MEGA labeled tryptic digest of proteins from membrane fraction of *E. coli* cell extract......263

Table 7.5 List of all proteins identified in *E. coli* membrane protein fraction......271

List of Abbreviations

MS	Mass spectrometry
ESI	Electrospray ionization
MALDI	Matrix-assisted laser desorption and ionization
LC	Liquid chromatography
DHB	2,5-Dihydroxybenzoic acid
HCCA	α-Cyano-4-hydroxycinnamic acid
TOF	Time-of-flight
MS/MS	Tandem mass spectrometry
PSD	Post-source decay
dc	Direct current
ac	Alternating current
rf	Radio frequency
m/z	Mass-to-charge ratio
CAD	Collision-activated dissociation
QTOF	Quadrupole time-of-flight
IT	Ion trap
RP	Reversed phase
2D	Two-dimentional
PAGE	Polyacrylamide gel electrophoresis
HPLC	High performance liquid chromatography
DIGE	Difference gel electrophoresis
3D	Three-dimensional
1D	One dimensional
PMF	Peptide mass fingerprinting
ppm	Parts per million
fmole	Fentomole
CID	Collision-induced dissociation
PTMs	Post-translational modifications
ICAT	Isotope-coded affinity tag

SILAC	Stable isotope labeling by amino acids in cell culture
AACM	Amino acid coded mass tagging
MCAT	Mass-coded abundance tagging
ACN	Acetoniltrile
TFA	Trifluoroacetic acid
Å	Angstrom
h	Hour
min	Minute
fmole	Fentomole
SDS	Sodium dodecylsulfate
DTT	Dithiothreitol
QqTOF	Quadrupole time-of-flight
SRS	Sequence retrieval system
OMPA	Outer membrane protein A
S/N	Signal-to-noise ratio
μLC	Microcapillary liquid chromatography
SCC9	Squamous carcinoma cells
SCC9-E	SCC9 cells expressing E-cadherin
ATTC	American type culture collection
MEM	Minimum essential medium
FBS	Fetal bovine serum
SCX	Strong cation exchange
PPD	Paraphenylene diamine
2MEGA	N-terminal dimethylation after lysine guanidination
SPE	Solid phase extraction
Leu-enk	Leucine enkephalin
m/z	Mass-to-charge ratio
E. coli	*Escherichia coli*
DVS	Divinyl sulfate
GRAVY	Grand average of hydropathy

Chapter 1

Introduction:

Modern Mass Spectrometry Combined with Separation Techniques for Qualitative and Quantitative Proteomics

The term proteome was first coined in 1994 to describe the set of proteins encoded by the genome [1]. Proteomics, the study of the proteome, has come to encompass the identification, characterization and quantification of the complete set of proteins expressed by the entire genome in the lifetime of a given cell, tissue or an organism, including isoforms, polymorphisms and modifications, protein-protein interactions and the structural description of proteins and their complexes. To date, mass spectrometry (MS) has become an indispensable analytical tool in proteomics, mainly due to the advent of the two 'soft ionization' techniques of electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) in the late 1980s. In the subsequent sections an overview is given of modern mass spectrometry-based methodologies for qualitative and quantitative proteomics. As well, a brief summary of the separation techniques used to simplify the complexity of biological samples will be outlined.

1.1 Ionization Methods

To analyze a sample by MS, it must be vaporized and ionized. The two ionization techniques most commonly used for the mass spectrometric analysis of proteins and peptides are ESI and MALDI. ESI produces gaseous ions from solution phase samples, and can therefore be easily coupled to liquid-based separation technologies, such as liquid chromatography (LC) and capillary electrophoresis. MALDI ionizes samples out of a dry, crystalline matrix, and can be used directly to analyze simple peptide mixtures as well as analyze complex peptide mixtures when coupled with offline LC.

1.1.1 MALDI

Matrix-assisted laser desorption ionization (MALDI), developed by two independent groups (Karas & Hillenkamp [2, 3], and Tanaka & coworkers [4, 5]) in the late 1980s, is one of the two most popular "soft ionization" methods. As shown in Figure 1.1, to generate gas phase, protonated molecules, analyte molecules (that is, the molecules to be analyzed) are co-crystallized with a large excess of matrix material by pipetting a submicroliter volume of the mixture onto a metal substrate and allowing it to dry. The resulting solid is then irradiated by nanosecond laser pulses, usually from a small nitrogen laser with a wavelength of 337 nm. The matrix is typically a small organic molecule that absorbs strongly at the wavelength of the laser employed. Figure 1.2 shows the chemical structure of two common matrices, 2,5-dihydrobenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (HCCA), that are almost exclusively used in biomolecule analysis. Matrices differ in the amount of energy they impart to the

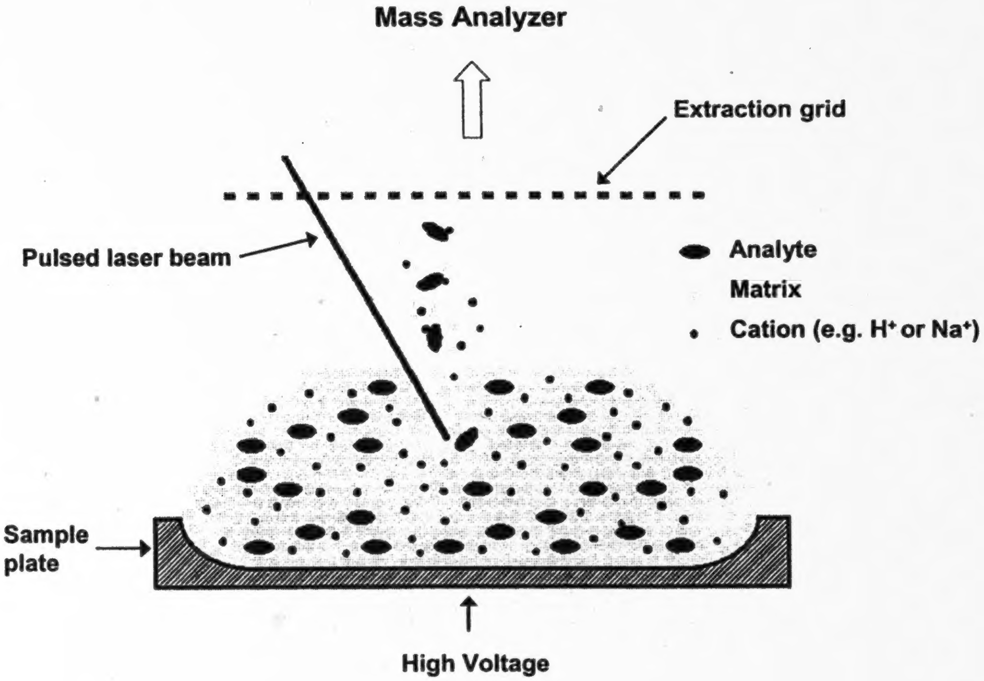
Mass Analyzer
Extraction grid
Pulsed laser beam
Analyte
Matrix
Cation (e.g. H⁺ or Na⁺)
Sample plate
High Voltage

Figure 1.1 Schematic diagram of MALDI process.

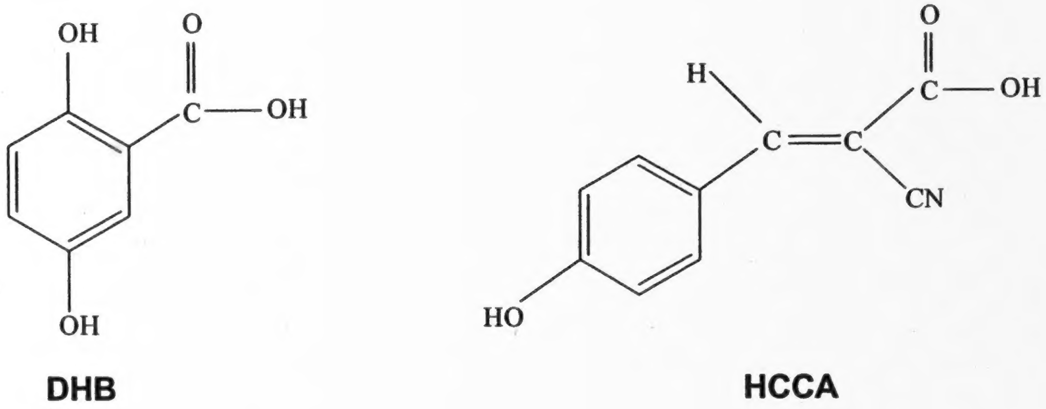
OH
O
C
OH
OH
DHB
O
H
C
OH
C
C
CN
HO
HCCA

Figure 1.2 Chemical structure of two common MALDI matrices.

biomolecules during desorption and ionization and hence the degree of fragmentation (unimolecular decay) that they cause. The HCCA matrix, which generally leads to the highest sensitivity in MALDI, results in "hotter" ions than the DHB matrix. The latter is therefore preferred when the ions need to be stable for milliseconds in trapping experiments rather than microseconds in time-of-flight experiments.

The formation of singly-protonated analytes in MALDI is typical, but the precise nature of the ionization process in MALDI is still largely unknown. The ion signal intensities depend on the level of incorporation of the peptides into crystals, their likelihood of capturing and/or retaining a proton during the desorption process, and a number of other factors, including suppression effects in peptide mixtures. For example, peptides with a C-terminal arginine generally generate higher signals than peptides with a C-terminal lysine [6]. For these reasons, it is difficult to relate peptide peak height with the quantity of sample present, unless an internal standard is used.

In contrast to ESI, MALDI has the following favorable attributes. First, due to the pulsed nature of most lasers, ions are formed in discrete events. If mass analysis is then synchronized with ion formation, very little sample is wasted. Therefore, MALDI can achieve very high levels of sensitivity, often providing data from sub-femtomole ($< 1\times10^{-15}$ mole) amounts of sample loading. A second advantage of MALDI lies in the fact that singly charged analytes are usually generated. When coupled with certain mass analyzers (e.g. time-of-flight (TOF)), MALDI can be used to provide molecular weight information rapidly for one or more analytes. High throughput can then be readily accomplished by using sample plates that are loaded with ~100 different samples. The third advantage of MALDI is that MALDI data acquisition can be stopped at any point and unconsumed sample recovered for later analysis. A final practical advantage of MALDI is its relatively high tolerance to salts and buffers.

Although MALDI has significant advantages, it also has some drawbacks. Even though the pulsed nature of the technique is one source of MALDI's inherent sensitivity, it is also a source of difficulty when coupling to some mass analyzers. Consequently, only certain mass spectrometers are easily coupled with MALDI. Also, the presence of a matrix, which facilitates ionization, causes a large degree of chemical noise to be observed at *m/z* ratios below 500 Da. As a result, samples with low molecular weights are usually difficult to be analyzed by MALDI.

1.1.2 Electrospray ionization

ESI MS has been developed for use in biological mass spectrometry by Fenn et al [7]. The ESI process (Figure 1.3) transfers ions in solution into gaseous ions at atmospheric pressure, which are sampled into the vacuum system of the mass spectrometer through a series of sampling apertures separating successive vacuum stages.

The mechanisms involved in the production of isolated gaseous ions by ESI are not fully understood [8]. The sample solution flows at low flow rates (nL/min to mL/min) through a capillary tube to which a high voltage (1–6 kV) is applied. Small charged droplets are sprayed from the ES capillary into a bath gas at atmospheric pressure and travel down a pressure and potential gradient towards an orifice in the mass-spectrometer high-vacuum system. As the droplets traverse this path they become desolvated and reduced in size to such an extent that surface-coulombic forces overcome surface-tension forces and the droplets break up into smaller droplets [8, 9]. This process continues until the point is reached that either an ion desorbs from a droplet [10]

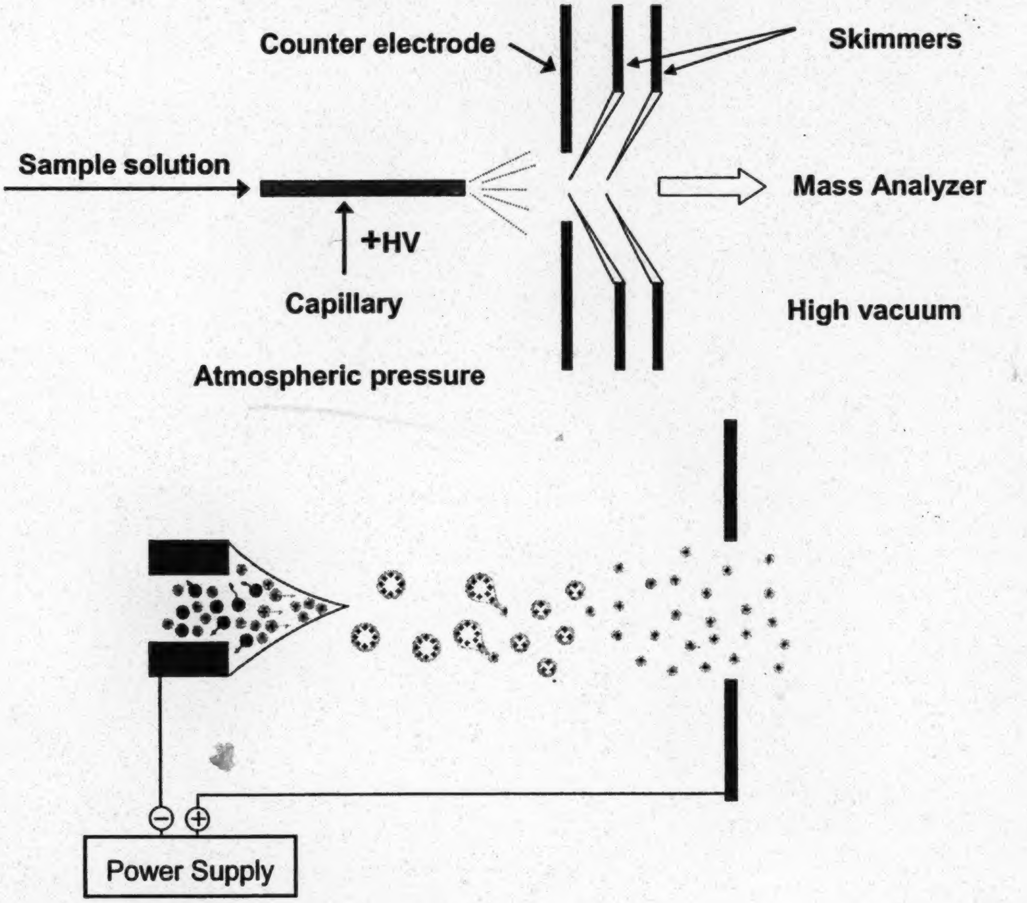
Counter electrode
Skimmers
Sample solution
Mass Analyzer
+HV
Capillary
High vacuum
Atmospheric pressure
⊖ ⊕
Power Supply

Figure 1.3 Schematic diagram of ES process

or solvent is completely removed [11]. The exact mechanism of ion formation, whether it is by ion evaporation (ion-evaporation model) or by complete solvent removal (charge-residue model), from the charged droplet is under debate, and in all probability different mechanisms apply in different situations [12-15]. Whatever the mechanism, the result is a beam of ions that are sampled by the mass spectrometer. Significantly, the ES process occurs at relatively low temperatures (room temperature, or just above), and so large, thermally labile, polar molecules can be ionized without decomposition. Ionized molecules of the form $[M + H]^+$ or $[M + nH]^{n+}$ (or $[M - H]^-/[M - nH]^{n-}$) are generally produced. The prerequisite for gaseous ion production with ES is that the analyte can be ionized in solution. If several ionizable sites are present, multiply charged ions will be produced; for example, denatured proteins typically carry one charge per 1000 Da [16]. By observing such multiply charged species, the effective mass range of the spectrometer can be extended to hundreds of thousands of Daltons.

Electrospray is typically performed in either the infusion mode, the nanoelectrospray format, or in combination with high-performance liquid chromatography. In the infusion mode, the sample is simply introduced into a continuous liquid stream (typically a mixture of organic and aqueous liquid such as 50:50 $MeOH:H_2O$) via an injection valve. Flow rates are usually between 0.5 and several microliters per minute. Samples have to be substantially free of salt and detergent, but can conveniently be cleaned up in a reversed-phase packing loop in the injector valve.

Nanoelectrospray [17, 18] is a miniaturized version of electrospray that operates without pumps and at very low flow rates in the range of a few nanoliters per minute. It is performed in pulled glass capillaries with an inner diameter at the tip of about one micrometer. A microliter volume of sample can be analyzed for more than an hour at full signal strength, which allows complex sequencing experiments to be performed. When liquid chromatography and mass spectrometry are coupled (LC-MS), MS analysis of the components of the sample takes place on-line as they elute from the chromatography column. In this case, sample cleanup, separation, and concentration are all achieved in a single step.

1.2 Mass Analyzers

1.2.1 TOF

Analysis by TOF is based on the following principles: an accelerating potential (V) will give an ion of charge *z* an energy of zV, which can be equated to the kinetic energy of the ion:

$$\mathbf{zV = 1/2mv^2}$$

where m = mass, v = velocity

If all ions are accelerated with the same potential, ions of different mass with the same charge must be traveling at different velocities. But velocity (v) = distance (d)/time (t), and therefore the equation can be rewritten:

$$m/z = 2V(t/d)^2$$

and ions with different mass will take different times to travel the same distance.

Mass-to-charge ratios are determined by measuring the time that ions take to move through a field-free region between the source and detector. The high sensitivity of TOF analyzers is attributable to high ion transmission, since all the m/z values in the flight tube can be accelerated. Mass resolution is affected by slight variations in flight time, and factors that create a distribution in flight times among ions with the same m/z ratio will result in poor mass resolution. Two techniques are used to compensate for temporal (time of ion formation), spatial (location of ion formation) and kinetic (energy of ion formation) distribution. The first technique is the use of ion mirrors, or reflectrons [19] (as shown in Figure 1.4), which compensate for variations in energy distribution. The reflectron creates a retarding field that deflects the ions, sending them back through the flight tube. The more energetic the ions, the deeper they penetrate the retarding field of the reflectron before being reflected. Thus a more energetic ion will travel a longer flight path and arrive at the detector at the time as less energetic ions of the same mass. The second technique is called time-lag focusing or delayed extraction [20-23]. By introducing a time delay between ion formation and extraction of ions from the source, wide spatial and temporal distribution can be avoided or greatly reduced.

MALDI is usually coupled to a TOF analyzer that measures the mass of peptides or intact proteins. Tandem mass spectrometry (MS/MS) can be achieved with reflectron TOF analyzers by the observation of post-source decay (PSD) fragments; however, this

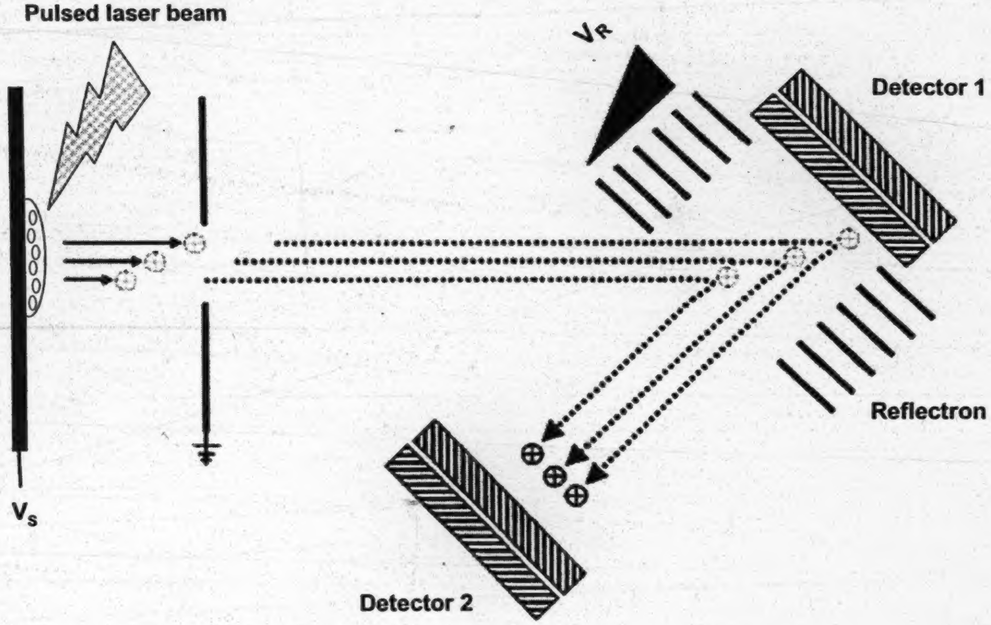

Figure 1.4 Schematic diagram of a reflectron MALDI TOF.

approach is protracted and arduous. MS/MS with TOF analysis is most commonly practiced by placing a collision cell between two TOF analyzers, or by configuring a TOF analyzer as the second stage in hybrid instruments (see below).

1.2.2 Quadrupole

Quadrupole mass analyzers consist of four precisely parallel rods equally spaced around a central axis. Opposing sets of rods have both a dc (direct current) and an ac (alternating current) or rf (radio frequency) voltage component, one set positive and the other set negative. Ions are introduced in a continuous beam along the central axis between the poles (Figure 1.5), and are filtered on the basis of their *m/z* ratios in the following manner: ions that pass between the two positive rods that are above a critical *m/z* ratio are transmitted through the centre of the quadrupole. This forms a high-pass mass filter. Ions that pass between the two rods with a negative potential that are below a critical *m/z* ratio are transmitted through the centre of the quadrupole; this forms a low pass mass filter. Combining both sets of rods into a quadrupole arrangement overlaps the two mass filter regions, creating a 'band pass' area of mutual stability (Figure 1.6) and allowing ions of a certain *m/z* ratio to pass through. Ions with *m/z* ratios outside this area of mutual stability cannot pass through and run into the rods. The *m/z* ratio of the ions that are allowed to pass through the quadrupole is proportional to the voltage applied to the rods; the higher the voltage, the higher the *m/z* value that is allowed to pass. By altering the relative contributions of the dc and rf components, the width of the band pass area, and therefore the resolution, can be adjusted (wider band pass = wider peak = lower resolution; narrower band pass = narrower peak = higher resolution). Scanning a

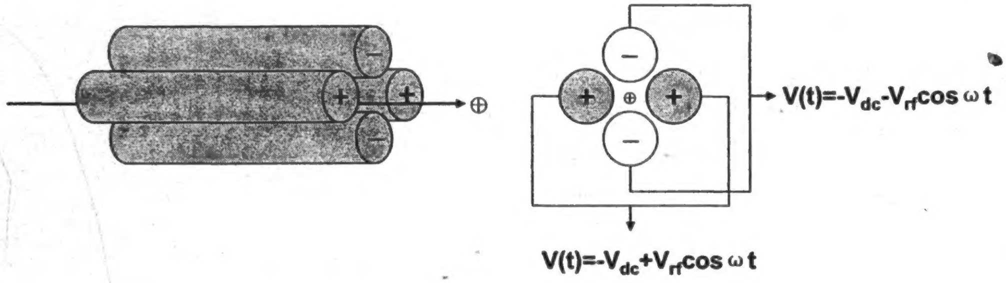
−
+
+
−
⊕
−
+
⊕
+
−
$V(t)=-V_{dc}-V_{rf}\cos\omega t$
$V(t)=-V_{dc}+V_{rf}\cos\omega t$

Figure 1.5 Schematic diagram of a quadrupole mass analyzer.

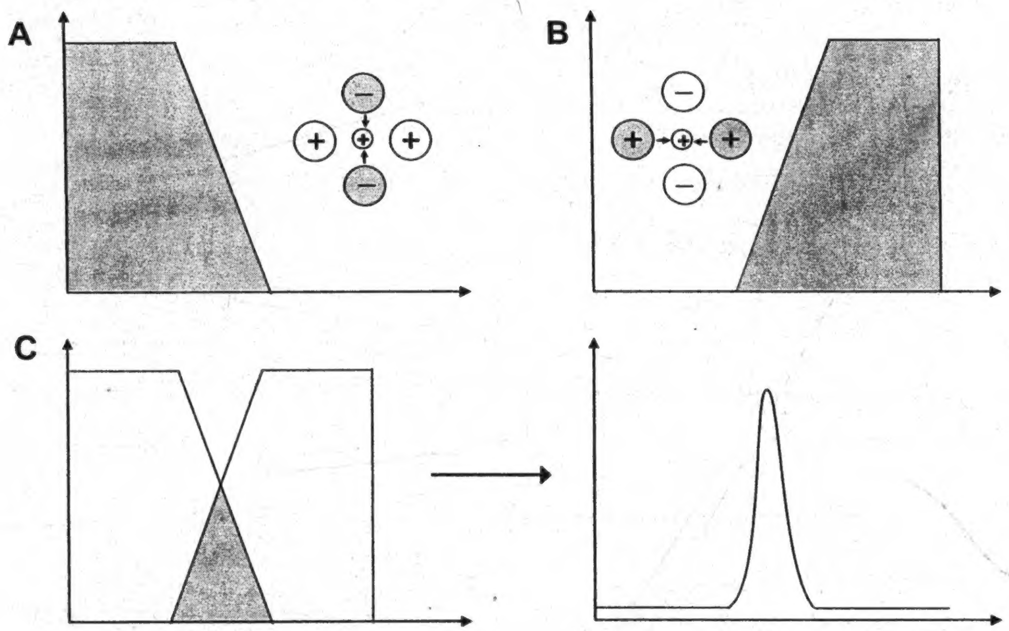

Figure 1.6 Mass filter created by a quadrupole mass analyzer. (A) Negative rods create a low-pass mass filter. (B) Positive rods create a high-pass mass filter. (C) Combining negative and positive sets of rods into a quadrupole arrangement overlaps the two mass filter regions, creating an area of mutual stability which allows ions of a certain m/z to pass.

quadrupole mass analyzer involves ramping the amplitude of the dc and rf voltages at a constant ratio, thus changing the position of the band pass region and allowing different masses to be transmitted.

For MS/MS analysis, three quadrupoles can be configured together (to form a 'triplequad'). The first and third quadrupoles are used for scanning, whilst the middle quadrupole is used as a collision cell. Ions in the second quadrupole are fragmented by collision-activated dissociation (CAD): low-energy collisions with an inert gas such as nitrogen or argon.

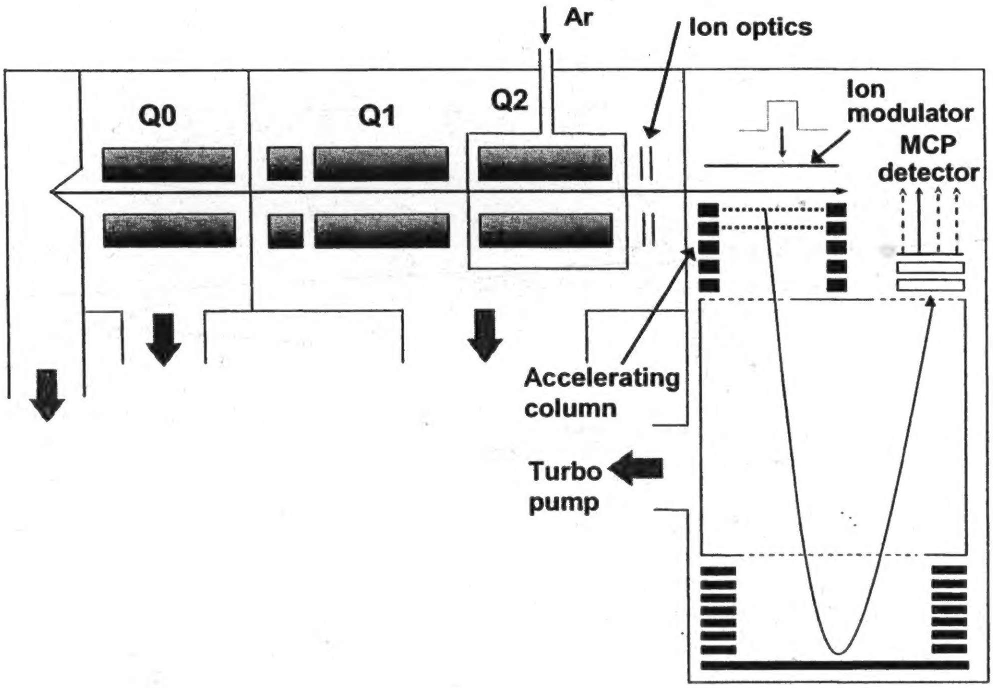

Figure 1.7 Schematic diagram of the tandem QTOF mass spectrometer.

About ten years ago, hybrid QTOF instruments were developed and commercialized, which can be described in the simplest way as a triple quadrupole with the last quadrupole section replaced by a TOF analyzer [24, 25]. Therefore, it is also known as QqTOF, where Q refers to a mass-resolving quadrupole, q refers to an r.f.-only quadrupole or hexapole collision cell and TOF refers to a time-of-flight mass spectrometer. In the usual QqTOF configuration, an additional r.f. quadrupole Q0 is added to provide collisional damping, so the instrument (Figure 1.7) consists of three quadupoles, Q0, Q1 and Q2, followed by a reflecting TOF mass analyzer with orthogonal injection of ions. The main advantages of these hybrid instruments are that they provide high mass accuracy and high resolution, resulting in unambiguous determination of charge state and very high specificity in database searches.

1.2.3 Ion Trap

In an ion trap (IT) analyzer, the ions are first captured or "trapped" in a three-dimensional electric field for a certain time interval and are then subjected to MS or MS/MS analysis (see Figure 1.8). The IT captures the continuous beam of ions up to the limit of their space charge [26, 27]. This is the maximum number of ions that can be introduced into the instrument without distorting the applied field. The ions are then subjected to additional electric fields, which eject one ion species after another from the

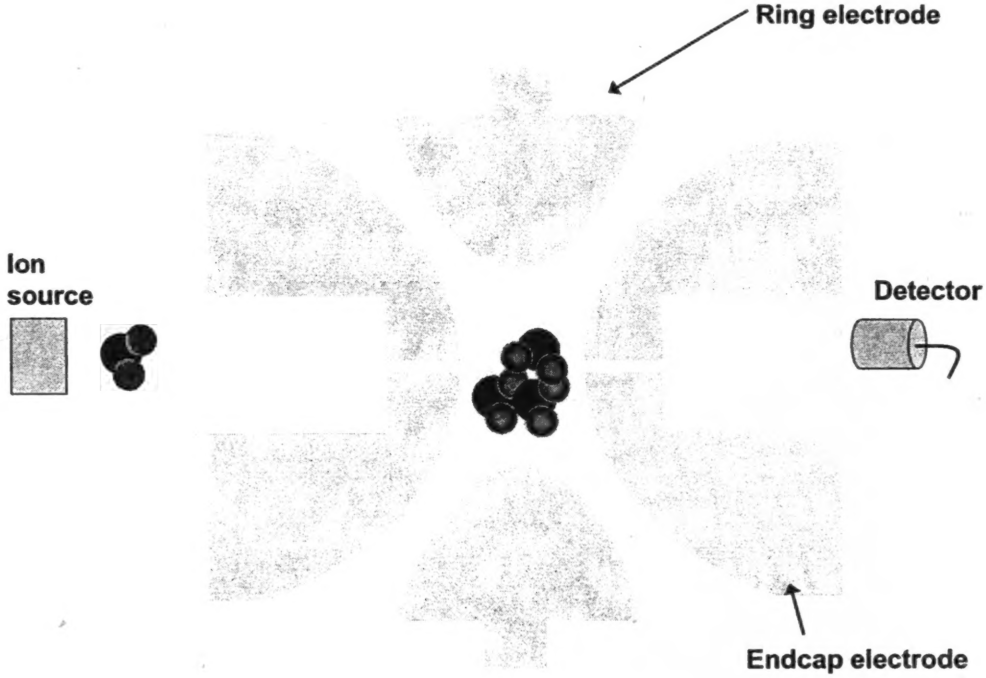

Figure 1.8 Schematic diagram of an ion trap mass analyzer.

trap, and are detected, to produce a mass spectrum. For an MS/MS experiment, all except the single desired ion species are ejected first. Then the remaining ion species is fragmented and its products are analyzed. Several sequential steps of MS/MS can in principle be performed, allowing detailed fragmentation studies. Owing to the operating principle of the trap, the lower end of the fragment mass range (below about a quarter of the parent ion mass) cannot be observed. Therefore, it is very rare to see immonium ions or a_1 ions in MS/MS spectra collected by IT instruments. The main advantages of IT are that it is automated, sensitive and relatively inexpensive. Therefore, coupled with RPLC and ES, ITs have been used to produce much of the proteomics data in the literature and are currently still in use. A disadvantage of the ITs is their relatively low mass accuracy, due in part to the limited number of ions that can be accumulated at their point-like centre before space-charging distorts their distribution and thus the accuracy of the mass measurement. Therefore, currently it is being replaced by high accuracy instruments, such as QTOF.

1.3 Separation Techniques

Proteomics represents a significant challenge to separation scientists because of the diversity and heterogeneity of proteins and peptides present in complicated biological systems. To reduce the sample complexity prior to MS, separation techniques such as two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE) or high performance liquid chromatography (HPLC) are recognized as indispensable tools in proteomics research. To date, mainly one of the following two approaches is usually taken prior to qualitative and quantitative proteomic analysis. One is called "top down" proteomics [28], in which proteins are initially separated, then digested and analyzed by MS; the other is called "shotgun" proteomics [29], in which a complex protein mixture is first digested and the resultant peptides are then chromatographically resolved and analyzed by MS. In both cases, separation technologies play a critical role in protein identification and analysis.

2D PAGE has been widely used to separate and to quantify intact proteins [30, 31] since thousands of proteins can be separated in a single analytical run. 2D PAGE offers the advantage that radioactive or fluorescence tagging can be used for the detection of post-translational modifications and for measurement of differential protein expression in cellular populations that differ in their physiological, metabolic or disease states [32]. Recently, two-color fluorescent labeling techniques, analogous to those used for detection of mRNA levels in microarray format, termed as 2D difference gel electrophoresis (DIGE), have been introduced for differential quantification in 2D PAGE [33]. 2D PAGE analysis has enabled the identification of many proteins from a vast variety of sources.

However, it has been recognized that 2D PAGE-based methods also suffer from the following technical limitations. First, although advances have been made, reproducibility remains a concern. Second, highly acidic and basic proteins and hydrophobic proteins are generally difficult to detect in 2D-PAGE separations. Third, sensitivity and dynamic range, i.e. the absolute amount of proteins that can be loaded onto the gel, are limited [34]. Therefore, low abundance proteins are usually not detected on 2D-PAGE. It was recently estimated on the basis of codon-bias distribution that more than half of all proteins in the yeast proteome are not detectable by 2D-PAGE analysis [35].

Because of its high resolution, speed, sensitivity, reproducibility and its compatibility with ESI and MALDI MS, HPLC represents an attractive alternative to 2D PAGE for the separation of both proteins and peptides. Compared with gel-based separation methods, LC-based separation methods minimize sample handling and preparation. Another attractive feature of liquid chromatography (LC) is the broad selection of stationary and mobile phases, which makes LC a versatile and fundamentally important tool in proteomics. During the last few years a number of LC techniques have appeared in the literature and a variety of modes have been used both alone and in combination to analyze proteins, among which the most popular at present are two-dimensional (2D): strong cation exchange/reversed-phase [29, 36] or three-dimensional (3D): strong cation exchange/avidin/reversed-phase [37] chromatographic separations of tryptic peptide mixtures resulting from protein samples that are frequently pre-fractionated by one-dimensional (1D) PAGE. It is possible to detect low-abundance proteins or peptides present in complex sample mixtures within a wide dynamic range of concentrations using those methods [30, 38], although considerable effort is required and a sufficient amount of starting protein sample must be available. Reversed-phase (RP)-HPLC has been primarily used as the last step of LC separations prior to MS since protein or peptide digests separated by RP-HPLC can be introduced directly into the mass spectrometer through ESI or collected onto MALDI plates using a variety of LC-MALDI interfaces for identification and analysis.

1.4 Protein Identification

1.4.1 Peptide Mass Fingerprinting

Protein identification can be performed using peptide mass fingerprinting (PMF), also referred to as peptide mass mapping. The standard approach to identify proteins includes separation of proteins by 1D, 2D PAGE or LC. Subsequently, the proteins are usually digested by a protease of high specificity, most commonly trypsin, which cleaves polypeptides at the C-terminal side of lysine and arginine residues, unless the next amino acid in the sequence is proline. Following digestion, MS analysis of the resulting mixture of peptides yields a peptide mass fingerprint, which is a set of measured molecular masses of proteolytic peptide ions obtained from the protein digest. In a database search, these mass values are matched against sets of proteolytic peptide masses calculated for all protein sequences in a sequence database. The search returns a list of the database entries with the highest number of matching peptide mass values, and various algorithms are used to rank the sequences and determine the probability that the highest ranking sequence entry is the true identity of the analyzed protein.

MALDI TOF MS is the most commonly used technique to perform PMF [39-42], because MALDI TOF MS is fast, robust, easy to perform, sensitive (low fmole range), accurate (low ppm range), and tolerant to a certain level of various contaminations such as salts and buffers. PMF can also be used to identify proteins in ESI spectra but it is seldom used because the peptide masses would need to be deconvoluted for each search [43]. In MALDI TOF mass spectrometry, peptides appear as singly charged species in the mass spectrum and this type of spectrum is simple to interpret, unlike an ESI mass spectrum which displays multiply charged species. Although PMF is an effective tool for the identification of relatively pure proteins, it fails to identify protein mixtures very often. Separation of complex protein mixtures by high-resolution 2D PAGE is well adapted to protein identification with PMF. On the other hand, the application of PMF in conjunction with either 1D PAGE or LC separations must be adjusted to the separation capacity.

Although PMF is relatively simple and a standard procedure can be described, various factors influence the outcome of the analysis. PMF identification relies on high mass accuracy [44] and to a greater extent, enzyme specificity. Without a known enzyme specificity PMF identification fails. PMF identification also relies on observing a large number of peptides when high mass accuracy is achieved. As a rule, at least five peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered for an unambiguous identification [45]. In many cases, PMF alone is not sufficient for unambiguous identification of a protein. This is particularly true when dealing with protein mixtures or small amounts of protein resulting in only a few peptides (less than five) that can be detected because of the MS detection sensitivity. In such cases, additional information, such as peptide sequence information, is required for a confident identification.

1.4.2 Database Searching with Tandem Mass Spectrometric Data

Although PMF works well in many cases, proteins can also be identified by interpreting the data resulting from fragmenting the peptides in tandem mass spectrometers [46, 47], which is a more specific and sensitive identification method. Tandem mass spectrometers and, to a more limited extent, single-stage mass spectrometers have the ability to fragment the selected peptide ions and to record the resulting fragment ion spectra. Analogously to viewing a PMF as a fingerprint of the analyzed protein, a fragment ion spectrum can be regarded as a fingerprint of the fragmented peptides, and thus be used directly for protein identification in sequence databases. In this technique, for each protein sequence in the database, the fragment ion masses recorded for one selected precursor peptide molecular ion are compared with the sets of possible fragment ion masses calculated for each proteolytic peptide sequence which matches the mass of the selected peptide ion.

For tandem mass spectrometers, such as triple quadrupole, ion trap, or QTOF instruments, fragment ion spectra are generated by a process called collision-induced dissociation (CID) in which the peptide ion to be analyzed is isolated and fragmented in a collision cell by collision with an inert gas, such as argon or nitrogen, and the fragment ion spectrum is recorded. Usually, these types of mass spectrometers are used in conjunction with ESI. Recently, MALDI combined with QqTOF has been used to generate MS/MS spectra for large-scale protein identification. Generally, low-collision energy is used to generate the tandem mass spectra.

Tandem mass spectra generated by the fragmentation of peptide ions in the gas phase at low collision energy are dominated by fragment ions resulting from cleavage at the amide bonds. Very little amino acid side chain fragmentation is observed. As shown in Figure 1.9, several bonds along the backbone can be broken by the collisions. The most common ion types are the b and the y ions, which denote fragmentation at the amide bond with charge retention on the N or C terminus, respectively. The nomenclature differentiates fragment ions according to where around the amide bond fragmentation occurs and the end of the peptide that retains a charge after fragmentation [48, 49]. If the positive charge associated with the parent peptide ion remains on the amino-terminal side of the fragmented amide bond, then this fragment ion is referred to as a *b* ion. However, the fragment ion is referred to as a *y* ion if the charge remains on the carboxyl-terminal side of the broken amide bond. Since in principle every peptide bond can fragment to generate a *b* or *y* ion, respectively, subscripts are used to designate the specific amide bond that was fragmented to generate the observed fragment ions: *b* ions are designated by a subscript that reflects the number of amino acid residues present on the fragment ion counted from the amino-terminus, whereas the subscript of *y* ions indicates the number of amino acids present, counting from the carboxyl-terminus.

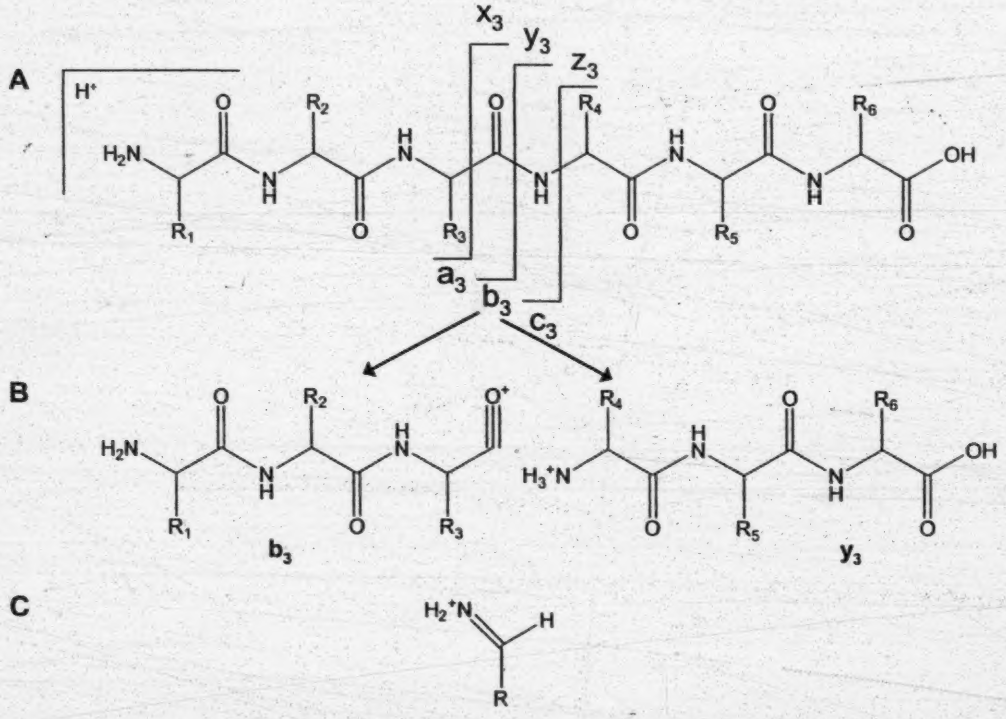

Figure 1.9 Peptide fragmentation nomenclature. (A) Nomenclature for peptide fragment ions that form via cleavage of bonds along the peptide backbone. (B) Example structure for b and y ions. (C) Immonium ion. Amino acids generate immonium ions to differing extents and some not at all.

The CID spectrum of a peptide ion acquired at low collision energy can be considered a composite of many discrete fragmentation events. Each peptide tandem mass spectrum will contain *b* and *y* ions as well as other fragment ions that can be used to interpret the amino acid sequence. These include diagnostic ions generated by the neutral loss of specific groups from amino acid side chains (e.g., the loss of ammonia (-17 u) from Gln, Lys, and Arg or of water (-18 u) from Ser, Thr, Asp and Glu) and low mass ions that result from the fragmentation of amino acids down to a basic unit consisting of the side chain residue and an immonium functionality (Figure 1.). The *b* ion series also often show a satellite ion series in which each signal is 28 u lower than the corresponding *b* ion. These signals result from the neutral loss of carbon monoxide and are referred to as an *a* ion series. CID spectra can be further complicated by the presence of internal fragment ions that represent some contiguous sequence of amino acids in the peptide. These are generated if a specific peptide ion undergoes two or more fragmentation events. Empirical observation shows that internal fragments often occur if either proline [50, 51] or aspartic acid [52] residues are present in a sequence and even more so at any aspartyl-proline bond [53], indicating that not all peptide bonds have the same propensity to fragment during low-energy CID. For the same reason, the relative intensity of fragment ions in peptide CID spectra is uneven and somewhat unpredictable. Some of the rules that control peptide ion fragmentation in a collision cell have been determined [54-56]; many others remain to be studied. If a proline residue is present in a peptide sequence, the most intense ions in the CID spectrum will generally be due to fragmentation on the amino-terminal side of proline. This is thought to occur because the gas-phase basicity of the proline imide bond is greater than that for any of the amide bonds. Under the moving proton hypothesis for CID, the proton available for fragmentation is therefore statistically more likely to be at this imide bond than at an amide bond in the peptide [54]. Additionally, it is known that peptides that contain aspartic acid tend to fragment at the carboxyl-terminal adjacent amide bond. This observation may be due to the ability of the aspartic acid side chain to influence the gas-phase basicity of the adjacent carboxyl-terminal amide bond via formation of a transient six-membered ring between the carboxylic acid group of the Asp side chain and the nitrogen of the adjacent amide bond.

Thus, the quality of peptide tandem mass spectra is dependent on the sequence location of amino acids, amino acid side chain basicity, amino acid side chain structure, and charge state of the peptide ion fragmented. If proteins are completely digested with trypsin, then lysine or arginine residues will be present at the carboxyl-terminus of all peptides except for the C-terminal peptide of the original protein. A charge sequestered by lysine or arginine at the C-terminus tends to produce a more complete series of *y* ion fragments than will be generated by peptides produced by protein digestion with chymotrypsin or other proteases where lysine and arginine are distributed throughout the sequences rather than at the C-terminus. Additionally, $[M + 2H]^{2+}$ ions of peptides will produce tandem mass spectra of higher quality than those from either $[M + H]^{+}$ or $[M + 3H]^{3+}$ peptide ions. The $[M + 2H]^{2+}$ peptide ions fragmented under low-energy CID produce spectra, although there are exceptions such as when proline and/or histidine are internal to the peptide sequence, that contain $[M + H]^{+}$ fragment ions that are more readily interpreted than tandem mass spectra of $[M + 3H]^{3+}$ and higher charge states that produce multiply charged fragment ions.

1.5 Post-translational modifications

Post-translational modifications (PTMs) are chemical processing events that alter the properties of a protein after its translation, by proteolytic cleavage or by addition of a modifying group to one or more amino acid residues. PTMs include phosphorylation, glycosylation, acetylation, methylation, sulphation, disulphide bond formation, deamidation and ubiquitination [57]. Most eukaryotic proteins are post-translationally modified, and many of these PTMs are regulatory and reversible, most notably, protein phosphorylation, which is a dynamic process with complex kinetics involving several amino acids in a single protein, and which controls biological function through many different mechanisms [58].

The application of MS to the identification of PTMs ranges from the study of single, purified proteins through the search for one type of modification on all the proteins in a sample to scanning for all modifications on a proteome-wide scale. The complexities involved in identifying all the modifications even on a single protein mean that proteome-wide scanning is, at present, not comprehensive; nevertheless, large amounts of biologically useful information can be generated. To determine the sites of modification, maximum protein sequence coverage is desirable. For this purpose, peptide mass mapping using two or more different enzymes can be employed, for example trypsin, Asp-N and Glu-C. Protein modifications are then identified from the measured masses and fragmentation spectra using manual or computer-assisted interpretation [58].

Several approaches that attempt to address the low-stoichiometry and high-complexity problems associated with the analysis of PTMs, and protein phosphorylation in particular, involve the selective enrichment of modified proteins. These techniques are generally based on some form of affinity selection that is specific for the modification of interest, and that is used for the purification of modified proteins.

1.6 Quantitative proteomics

Beyond enumerating the proteins expressed in a species, the detection and quantification of differences in the protein profiles of cells, tissues or body fluids of different origins or states is increasingly being recognized as a key objective of proteomics research [59]. The measurement of differential protein expression provides a more direct, more accurate and more versatile way to detect global changes in cellular dynamics in health and disease than the complementary and more mature technology of mRNA expression analysis [60]. Quantitative or comparative proteomics thus holds significant promise for the discovery of diagnostic or prognostic protein markers, for the detection of new therapeutic targets and as a powerful tool to further our understanding of basic biological processes and mechanisms. The realization of these expectations relies on the development of automated, robust and highly sensitive methods to identify and quantify proteins.

Quantitative or comparative proteome analysis was initially performed with 2D PAGE with the inherent disadvantages of being biased towards certain proteins and being labor intensive. Alternative mass spectrometry-based approaches in conjunction with gel-free protein/peptide separation have been developed in recent years using various stable isotope labeling techniques. Common to all these techniques is the incorporation, biosynthetically or chemically, of a labeling moiety having either a natural isotope distribution of hydrogen, carbon, oxygen, or nitrogen (light form) or being enriched with heavy isotopes like deuterium, ^{13}C, ^{18}O, or ^{15}N, respectively. By mixing equal amounts of a control sample possessing for instance the light form of the label with a heavy-labeled case sample, differentially labeled peptides are detected by mass spectrometric methods and their intensities serve as a means for direct relative protein quantification.

1.6.1 The Principle

To detect peptides using mass spectrometry, the analyte molecules have to be ionized and transferred into the gas-phase. This ionization process depends on several factors such as the physico-chemical properties of the analyte molecules themselves (ionization efficiency) and the presence of other components, including buffer salts and other peptides, present in the sample mixture at the time of ionization (suppression effects). This implies that the intensity of a particular peptide signal is not simply a function of its abundance. Therefore, the ion signals of two different peptides (even when they originate from the same protein) cannot be compared with each other. Further, even the signal intensities of the same peptide ion obtained from two independent experiments cannot be compared. To overcome this limitation, it is essential to use an internal standard, which allows useful quantitative information to be generated. The most important prerequisite for an internal standard to be reliable is that it should be chemically as similar as possible to the analyte being analyzed. Only this ensures comparable ionization efficiencies and suppression effects. The best internal standard is a molecule that has an identical chemical structure but is labeled with stable isotopes such as ^{2}H, ^{13}C, ^{15}N and ^{18}O. Labeling with a stable isotope changes the mass of the analyte and is, therefore, the perfect combination with mass spectrometry because this technique can easily distinguish labeled and corresponding non-labeled analytes. Given that mass spectrometry coupled with liquid chromatography is usually used for quantification, the retention behavior of the analytes and their internal standards should also be as similar as feasible. Deuterium-labeling could have a significant effect on the retention behavior [61-63], therefore, ^{13}C, ^{15}N or ^{18}O-labeling might be the better choice.

The pioneer study by Chait and coworkers on ^{15}N-metabolic labeling [64] and that by Aebersold and coworkers on isotope-coded affinity tags (ICAT) [65], which were described in two seminal articles on methods for relative quantification, became the foundation for all subsequent studies describing approaches for the quantification of proteins by mass spectrometry. Over the last six years, many attempts have been made to develop strategies, based on stable isotope labeling and MS, for quantitative analysis of proteins between two samples. Normally, one sample is labeled with a heavy reagent and a second sample is labeled with a light reagent. The two samples are then mixed and analyzed by MS. The ratio between two isotopic distributions (one for the light reagent the other for the heavy reagent) can then be determined from the mass spectra and used to calculate the relative protein quantities. Generally, these strategies can be divided into three categories depending on the manner in which the labeling is accomplished (Figure 1.10).

1.6.2 Isotopic Labeling

1.6.2.1 In Vivo

The *in vivo* labeling (also named metabolic labeling) method was first reported by Oda *et al* [64], in which yeast cells were grown in two separate media: one contained light isotopes, and the other contained a heavy isotope, ^{15}N. The two yeast cultures were combined, the proteins extracted, fractionated and then separated by gel electrophoresis. Finally, the proteins of interest were digested with trypsin before MS analysis and the relative quantities determined from the isotopic distribution ratios. Metabolic

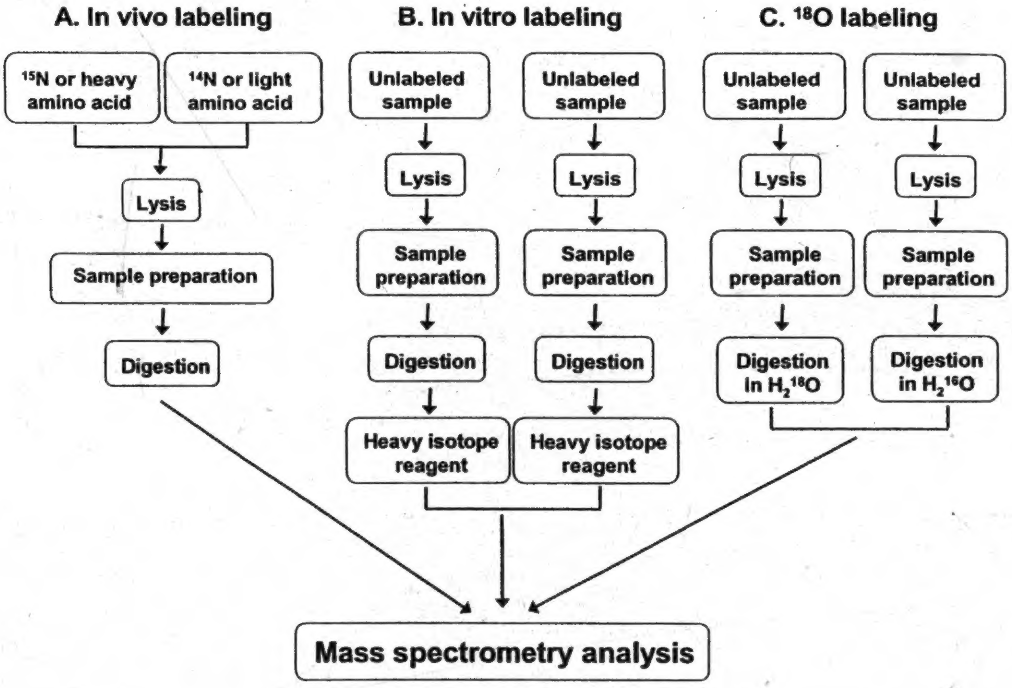
A. In vivo labeling
B. In vitro labeling
C. ¹⁸O labeling
¹⁵N or heavy amino acid
¹⁴N or light amino acid
Lysis
Sample preparation
Digestion
Unlabeled sample
Lysis
Sample preparation
Digestion
Heavy isotope reagent
Unlabeled sample
Lysis
Sample preparation
Digestion in H₂¹⁸O
Digestion in H₂¹⁶O
Mass spectrometry analysis

Figure 1.10 Schematic representation of different strategies for stable-isotope protein or peptide labeling for quantitative proteomics. Note: for the in vitro labeling strategy, some stable isotope reagents (e.g. ICAT) can label proteins instead of peptides.

incorporation of the stable isotope (^{15}N) labeled nutrients in growth media into cultured cells was used in a similar way for analyzing whole cell lysates from bacteria [66, 67].

Like the use of enriched/depleted ^{15}N and ^{13}C media, in vivo incorporation of isotopically coded amino acids has been a popular method to perform comparative proteomics. Selected amino acids are incorporated into proteins by growing cell lines that are auxotrophic in the chosen amino acids. One cell line is grown in normal media and another in media rich in an isotopically coded amino acid. After a period of time, the two cell lines are harvested and analyzed in the same manner as other quantification studies based on metabolic labeling.

Both essential and nonessential amino acids have been quantitatively incorporated into cell lines [68] These methods have come to be known by the acronyms SILAC [69, 70] (Stable Isotope Labeling by Amino acids in Cell culture) and AACM (Amino Acid Coded Mass tagging) [68]. The inherent advantage of this type of metabolic labeling is that it provides predictable mass shifts between peptide pairs. Moreover, the mass difference of differentially coded peptide pairs provides information on the number of amino acids in a peptide, thereby constraining database searches. Choice of the amino acid is crucial. Amino acid abundance will determine the number of tryptic peptides in mixtures of control and experimental samples that are differentially labeled. Lysine is a popular choice due to its relatively high abundance, the fact that proteases such as Lys-C and trypsin produce peptides with a single lysine residue, and the lack of metabolic scrambling of the label during protein synthesis [71]. Leucine is another abundant amino acid that gives broad labeling of a proteome [70, 72-74].

In vivo labeling has been proven to be an effective way for performing quantitative proteomic analysis since the internal standard is introduced at an early stage of the proteome study, thus avoiding the variations in downstream protein sample workup, which are normally associated with in vitro labeling and ^{18}O labeling. However, this approach can not be applied to tissues or body fluids and is limited to cells that can be grown in culture in 'controlled media'. The method also has several limitations. It is impossible to code globally all of the proteins in a proteome, not even with mixtures of labeled amino acids [75, 76]. The use of mixtures also leads to a sequence dependent mass shift, thereby complicating spectral interpretation. Metabolic scrambling and the concomitant uncertainty introduced in quantification is another issue. The amino acid being incorporated should be at the end of a metabolic pathway to avoid conversion to other amino acids.

1.6.2.2 In Vitro

In vitro labeling, also named chemical labeling, offers an excellent alternative quantitative tool when metabolic labeling is not suited to an analysis. The first chemical labeling method was the use of isotope-coded affinity tag (ICAT) reagents to quantitate cysteine-containing proteins between two samples [65]. The reagents comprise a cysteine-reacting group, a labeled linker containing either light or heavy isotopes, and an affinity group for separation. The main advantage of this method is that it enriches peptides containing the rare amino acid cysteine, thereby significantly reducing the complexity of the peptide mixture and increasing the dynamic range of MS analysis [77, 78]. However, early application of this method identified several drawbacks of the first-generation ICAT reagents. First, the eight deuterium atoms associated with this mass tag can lead to partial resolution of isotopic peptide pairs by HPLC, complicating the MS analysis [37, 63]. Second, the bulky affinity group, biotin, increases the complexity in the interpretation of MS/MS spectra. Third, ICAT is limited to Cys-containing proteins that do not cover the whole proteome. The first problem can be addressed by using a ^{13}C labeling strategy instead of deuterium labeling [63]. The second problem may be resolved by using a solid-phase capture-and-release system bearing a photocleavable linker that reduces the tag size before MS analysis [79], or by using an acid-labile linker [80].

Besides isotopic coding of the cysteine side chains using ICAT or similar reagents, a variety of chemical labeling reagents have been designed for the quantification of global protein expression. One class of this kind of reagents labels the N-termini and/or lysine side chains of tryptic peptides, such as acylation of the N-termini and ε-amino units of Lys residues by acetic anhydride ($-H_6$ and $-D_6$) or *N*-acetoxysuccinimide ($-H_3$ and $-D_3$) [81-83], succinylation of N-termini using 1-Nicotinoyloxy succinimide Esters ($-H_4$ and $-D_4$). Acylation of basic amino groups, however, changes the ionic states of peptides and may reduce the ionization efficiency of tryptic digests containing C-terminal lysine residues. Another class of global chemical labeling method is the labeling the carboxylic acid groups on the side chains of aspartic acid, glutamic acid and C-termini of peptides using methanol ($-H_3$ and $-D_3$) [84]. The small mass difference generated between peptide pairs when deuterated reagents are used limits the application of these labeling methods.

Several chemical labeling methods that do not use isotope reagents have also been reported for quantification of protein expression, including mass-coded abundance tagging (MCAT) labeling of lysine residues with O-methylisourea [85], and differential amidination of N-termini and lysine residues with either acetamidine ("methyl-coded") or propionamidine ("ethyl-coded") moieties [86]. Although these procedures are simple and lead to higher ionization efficiency, several issues related to the difference in the physicochemical characteristics between the labeled and unlabeled peptides markedly reduce the accuracy of the quantification.

1.6.2.3 Labeling during proteolysis (^{18}O labeling)

Incorporating ^{18}O or ^{16}O into the C-termini of peptides is one of most popular stable isotope labeling methods that have been widely used in quantitative proteome analysis. A hydroxyl group from water is introduced into the carboxyl group formed during proteolysis. Therefore, when proteolysis is carried out in $H_2{}^{18}O$, all peptides will be labeled except the peptide originating from the carboxy-terminus of the protein. When proteolysis of control and experimental samples is carried out in $H_2{}^{16}O$ and $H_2{}^{18}O$, respectively, the peptides are differentially coded according to sample origin and the ratios of differentially labeled peptide pairs have been used for the quantification of proteins [87, 88]. It is even possible to ^{18}O label peptides after proteolysis by a back exchange reaction in which $H_2{}^{16}O$ is removed and replaced with $H_2{}^{18}O$ [89, 90]. This enzymatic approach has also been used to label a peptide's C-terminus, thereby creating a recognizable signature for peptide fragments containing this group and thus helping in assigning the y-type of peptide fragments in spectra resulting in more reliable protein identification.

Although ^{18}O is a simple and useful isotopic labeling method, it also has some drawbacks. First, systematic studies have shown that proteolytic enzymes incorporate different levels of isotopes from water during proteolysis, which complicates the data analysis process. For example, Asp-N and chymotrypsin incorporate one ^{18}O atom into the C-termini of peptides, whereas trypsin, Lys-C, and Glu-C incorporate either one or two ^{18}O atoms [91]. Second, the small mass difference between peptide pairs is not large enough to eliminate the significant overlap of isotope envelopes, especially for those peptide pairs with m/z over 2000 Da. Therefore, additional deconvolution steps are required to get accurate quantification results. Third, with ^{18}O-based quantification of expression is the need to evaporate samples to dryness before the introduction of $H_2{}^{18}O$. Redissolution of peptides is sequence dependent and prone to losses, requiring special sample preparation methods [88]. Finally, there is also the problem that peptides from the C-terminus of proteins appear as singlets. These singlets can be interpreted as having arisen from large changes in expression. This limitation also makes the ^{18}O labeling difficult to analyze serum samples quantitatively since a high percentage of identified peptides in serum samples are nontryptic peptides [92]. The number of nontryptic peptides observed in serum is much greater than that observed for the analysis of cell lysates (e.g., cell culture or tissue) since products of protein breakdown and turnover are commonly present in serum [93, 94].

1.6.3 *De Novo* Sequencing

Beyond quantification, some of the presented MS-based quantification methods may additionally aid in *de novo* sequencing of peptides [85, 95]. Attachment of labels to the peptide termini may enhance or decrease the ionization rate of distinct ion series, thereby simplifying MS/MS spectra. Furthermore, certain ion series can be indicated either directly by isotopic patterns in MS/MS experiments or by comparison of spectra of corresponding light and heavy peptides. Therefore, the lightly labeled peptide (parent ion) is fragmented first, followed by fragmentation of the heavily labeled peptide in a second step. Subsequently, these two fragment ion spectra are overlaid. The resulting spectrum shows a doubled y- or b-ion series depending on the labeled peptide terminus. If the label is not attached C- or N-terminally, the doubled ion series switches from one ion series to the other at the specific amino acid to which the label is attached. Therefore, the amino acid sequence can be read directly.

1.7 Objectives of the thesis

The first objective of this work was to develop a method for characterization of PTMs of low-mass proteins by combining separation tools with intact protein mass measurement and protein digest analysis.

Most of my efforts in this work were focused on developing a simple, economic, efficient and reliable method for global quantitative proteomic analysis, based on stable isotopic labeling, multiple dimensional LC separations and microbore LC MALDI QqTOF, and applying the developed strategy to identify biologically interesting proteins found to be different in abundance in different human squamous carcinoma cells prepared under different conditions. The ability of the developed strategy to study post-translational modifications (PTMs) such as phosphorylation, and de novo sequencing was also demonstrated using mature bovine milk as a study model. In addition, three different levels of variations which correspond to different starting points for sample preparation, such as cell growth, protein extraction, and protein, will be investigated. Finally, the effectiveness of a newly developed method, N-terminal dimethyl labeling, on the large-scale proteome analysis will be evaluated.

1.8 Literature Cited

(1) Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F. *Biotechnology (N Y)* **1996**, *14*, 61-65.

(2) Karas, M.; Bachmann, D.; Bahr, U.; Hillenkamp, F. *International Journal of Mass Spectrometry and Ion Processes* **1987**, *78*, 53-68.

(3) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299-2301.

(4) Tanaka, K.; Ido, Y.; Akita, S. *In Proceedings of the Second Japan-China Joint Symposium on Mass Spectrometry*; Bando Press: Osaka, Japan, 1987.

(5) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.

(6) Krause, E.; Wenschuh, H.; Jungblut, P. R. *Anal. Chem.* **1999**, *71*, 4160-4165.

(7) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(8) Gaskell, S. J. *J. Mass Spectrom.* **1997**, *32*, 677-688.

(9) Kebarle, P.; Ho, Y. *On the mechanism of electrospray mass spectrometry*; John Wiley and Sons: New York, 1997.

(10) Thomson, B. A.; Iribarne, J. V. *J. Chem. Phys.* **1979**, *71*, 4451-4463.

(11) Dole, M.; Mach, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968**, *49*, 2240-2247.

(12) Tang, L.; Kebarle, P. *Anal. Chem.* **1991**, *63*, 2709-2715.

(13) Tang, L.; Kebarle, P. *Anal. Chem.* **1993**, *65*, 3654-3668.

(14) Kebarle, P. *J. Mass Spectrom.* **2000**, *35*, 804-817.

(15) Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65*, 972A-986A.

(16) Hoffmann, E. d.; Stroobant, V. *Mass Spectrometry: Principles and Applications*; Wiley: Chichester, 2003.

(17) Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1-8.

(18) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, *379*, 466-469.

(19) Mamyrin, B.; Karataev, V.; Shmikk, D.; Zagulin, V. *Sov. Phys. JETP* **1973**, *37*,

45-48.

(20) Vestal, M. L.; Juhasz, P.; Matin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044-1050.

(21) Colby, S. M.; King, T. B.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 865-868.

(22) Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 1998-2003.

(23) Whittal, R. M.; Schriemer, D. C.; Li, L. *Anal. Chem.* **1997**, *69*, 2734-2741.

(24) Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889-896.

(25) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015-1024.

(26) Louris, J. N.; Cooks, R. G.; Syka, J. E. P.; Kelley, P. E.; Stafford, G. C.; Todd, J. F. J. *Anal. Chem.* **1987**, *59*, 1677-1685.

(27) Jonscher, K. R.; Yates, J. R. *Anal. Biochem.* **1997**, *244*, 1-15.

(28) Kettman, J. R.; Frey, J. R.; Lefkovits, I. *Biomol. Eng.* **2001**, *18*, 207-212.

(29) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd *Anal. Chem.* **2001**, *73*, 5683-5690.

(30) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat. Biotechnol.* **2001**, *19*, 242-247.

(31) Pandey, A.; Mann, M. *Nature* **2000**, *405*, 837-846.

(32) Baty, J. W.; Hampton, M. B.; Winterbourn, C. C. *Proteomics* **2002**, *2*, 1261-1266.

(33) Unlu, M.; Morgan, M. E.; Minden, J. S. *Electrophoresis* **1997**, *18*, 2071-2077.

(34) Shen, Y.; Smith, R. D. *Electrophoresis* **2002**, *23*, 3106-3124.

(35) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9390-9395.

(36) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., 3rd *Nat. Biotechnol.* **1999**, *17*, 676-682.

(37) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946-951.

(38) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. *J. Proteome Res.* **2002**, *1*, 47-54.

(39) Pappin, D. J.; Hojrup, P.; Bleasby, A. J. *Curr. Biol.* **1993**, *3*, 327-332.

(40) Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338-345.

(41) James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58-64.

(42) Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5011-5015.

(43) Yates, J. R., 3rd; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397-408.

(44) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741-4750.

(45) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437-473.

(46) Biemann, K.; Scoble, H. A. *Science* **1987**, *237*, 992-998.

(47) Hunt, D. F.; Shabanowitz, J.; Yates, J. R., 3rd; Zhu, N. Z.; Russell, D. H.; Castro, M. E. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 620-623.

(48) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

(49) Biemann, K. *Methods Enzymol.* **1990**, *193*, 886-887.

(50) Schwartz, B. L.; Bursey, M. M. *Biol. Mass Spectrom.* **1992**, *21*, 92-96.

(51) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1993**, *65*, 425-438.

(52) Goodlett, D. R.; Gale, D. C.; Guiles, S.; Crowther, J. *Encyclopedia of Analytical Chemistry*; John Wiley: New York, 2000.

(53) Mak, M.; Mezo, G.; Skribanek, Z.; Hudecz, F. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 837-842.

(54) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365-8374.

(55) Harrison, A. G.; Csizmadia, I. G.; Tang, T. H. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 427-436.

(56) Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 6233-6237.

(57) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255-261.

(58) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.

(59) Tao, W. A.; Aebersold, R. *Curr. Opin. Biotechnol.* **2003**, *14*, 110-118.

(60) Duggan, D. J.; Bittner, M.; Chen, Y.; Meltzer, P.; Trent, J. M. *Nat. Genet.* **1999**, *21*, 10-14.

(61) Zhang, R.; Regnier, F. E. *J. Proteome Res.* **2002**, *1*, 139-147.

(62) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. *Anal. Chem.* **2002**, *74*, 3662-3669.

(63) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. *Anal. Chem.* **2001**, *73*, 5142-5149.

(64) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 6591-6596.

(65) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

(66) Smith, R. D.; Pasa-Tolic, L.; Lipton, M. S.; Jensen, P. K.; Anderson, G. A.; Shen, Y.; Conrads, T. P.; Udseth, H. R.; Harkewicz, R.; Belov, M. E.; Masselon, C.; Veenstra, T. D. *Electrophoresis* **2001**, *22*, 1652-1668.

(67) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., 3rd *Anal. Chem.* **2002**, *74*, 1650-1657.

(68) Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. *Anal. Chem.* **2003**, *75*, 1316-1324.

(69) Ong, S. E.; Kratchmarova, I.; Mann, M. *J. Proteome Res.* **2003**, *2*, 173-181.

(70) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell. Proteomics* **2002**, *1*, 376-386.

(71) Gu, S.; Pan, S.; Bradbury, E. M.; Chen, X. *Anal. Chem.* **2002**, *74*, 5774-5785.

(72) Jiang, H.; English, A. M. *J. Proteome Res.* **2002**, *1*, 345-350.

(73) Pratt, J. M.; Robertson, D. H.; Gaskell, S. J.; Riba-Garcia, I.; Hubbard, S. J.;

Sidhu, K.; Oliver, S. G.; Butler, P.; Hayes, A.; Petty, J.; Beynon, R. J. *Proteomics* **2002**, *2*, 157-163.

(74) Martinovic, S.; Veenstra, T. D.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. *J. Mass Spectrom.* **2002**, *37*, 99-107.

(75) Chen, X.; Smith, L. M.; Bradbury, E. M. *Anal. Chem.* **2000**, *72*, 1134-1143.

(76) Zhu, H.; Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2115-2123.

(77) Wang, S.; Regnier, F. E. *J. Chromatogr. A* **2001**, *924*, 345-357.

(78) Spahr, C. S.; Susin, S. A.; Bures, E. J.; Robinson, J. H.; Davis, M. T.; McGinley, M. D.; Kroemer, G.; Patterson, S. D. *Electrophoresis* **2000**, *21*, 1635-1650.

(79) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, *20*, 512-515.

(80) Qiu, Y.; Sousa, E. A.; Hewick, R. M.; Wang, J. H. *Anal. Chem.* **2002**, *74*, 4969-4979.

(81) Che, F. Y.; Fricker, L. D. *Anal. Chem.* **2002**, *74*, 3190-3198.

(82) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. *J. Chromatogr. B Biomed. Sci. Appl.* **2000**, *745*, 197-210.

(83) Geng, M.; Ji, J.; Regnier, F. E. *J. Chromatogr. A* **2000**, *870*, 295-313.

(84) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; von Haller, P.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1214-1221.

(85) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163-170.

(86) Beardsley, R. L.; Reilly, J. P. *J. Proteome Res.* **2003**, *2*, 15-21.

(87) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836-2842.

(88) Stewart, II; Thomson, T.; Figeys, D. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2456-2465.

(89) Yao, X.; Afonso, C.; Fenselau, C. *J. Proteome Res.* **2003**, *2*, 147-152.

(90) Staes, A.; Demol, H.; Van Damme, J.; Martens, L.; Vandekerckhove, J.; Gevaert, K. *J. Proteome Res.* **2004**, *3*, 786-791.

(91) Reynolds, K. J.; Yao, X.; Fenselau, C. *J. Proteome Res.* **2002**, *1*, 27-33.

(92) Xiao, Z.; Conrads, T. P.; Lucas, D. A.; Janini, G. M.; Schaefer, C. F.; Buetow, K. H.; Issaq, H. J.; Veenstra, T. D. *Electrophoresis* **2004**, *25*, 128-133.

(93) Wall, D. B.; Kachman, M. T.; Gong, S.; Hinderer, R.; Parus, S.; Misek, D. E.; Hanash, S. M.; Lubman, D. M. *Anal. Chem.* **2000**, *72*, 1099-1111.

(94) Kachman, M. T.; Wang, H.; Schwartz, D. R.; Cho, K. R.; Lubman, D. M. *Anal. Chem.* **2002**, *74*, 1779-1791.

(95) Hsu, J. L.; Huang, S. Y.; Shiea, J. T.; Huang, W. Y.; Chen, S. H. *J. Proteome Res.* **2005**, *4*, 101-108.

Chapter 2

A New Mass Spectrometry Method for the Study of Post-Translational Modifications of Low-Mass Proteins [a]

2.1 Introduction

Post-translational modifications (PTMs) of proteins are chemical processing events by proteolytic cleavage or by addition of a modifying group to one or more amino acid residues that alter a protein's properties after its translation. PTMs include phosphorylation, glycosylation, acetylation, methylation, sulfation, disulfide bond formation, deamidation, ubiquitination, and nitration of tyrosine [1]. Most proteins undergo co- and/or post-translational modifications. Knowledge of these modifications is extremely important because they may alter proteins' physical and chemical properties, folding, conformational distribution, stability, and activity, and consequently change the function of the proteins and their interaction with other proteins. Moreover, the modification itself can act as an added functional group. Examples of the biological effects of protein modifications include phosphorylation for signal transduction, ubiquitination for proteolysis, attachment of fatty acids for membrane anchoring and association. Consequently, the analysis of proteins and their post-translational modifications is particularly important for the study of heart disease, cancer, neurodegenerative diseases and diabetes.

---

[a] A portion of this chapter was presented at the 51st Annual ASMS Conference, Montreal, Canada, 2003 as: C. Ji, Z. Wang and L. Li, "A New Mass Spectrometry Method for the Study of Post-translational Modifications of Low-Mass Proteins".

Despite the major role of PTMs in biological functions, large-scale study still remains as a great challenge due to a lack of suitable methods [1]. Recent applications of MS to the identification of PTMs have ranged in scale from the study of single, purified proteins to the search for one type of modification on all the proteins in a sample, and to scanning for all modifications on a proteome-wide scale. Considering the complexities involved in identifying all the modifications on a single protein, proteome-wide scanning is, at present, not comprehensive. Nevertheless, large amounts of biologically useful information can be generated. To determine the sites of modification, maximum protein sequence coverage is desirable. For this purpose, peptide mass mapping using two or more different enzymes, such as trypsin, Asp-N or Glu-C, can be employed. Protein modifications are then identified from the measured masses and fragmentation spectra using manual or computer-assisted interpretation [2].

To date, analysis of PTMs has not been routinely carried out, in part, because of the inability of current methods to decipher the modifications in a biological mixture. This study presents a method for the characterization of PTMs of low-mass proteins (molecular weight (MW): 5~20K) by combining separation tools with intact protein mass measurement and protein digest analysis.

2.2 Experimental

2.2.1 Chemicals and Reagents

HPLC grade acetone, methanol, acetonitrile (ACN), 2-propanol, and glacial acetic acid were from Fisher Scientific Canada (Edmonton, AB, Canada). Water used in all experiments was obtained from a Milli-Q Plus purification system (Millipore Corp., Bedford, MA). Bovine trypsin, cytochrome c, trifluoroacetic acid (TFA) and α-cyano-4-hydroxycinnamic acid (HCCA) were from Sigma-Aldrich-Fluka (Oakville, ON, Canada). HCCA was purified by recrystallization from ethanol (95%) at 50 °C prior to use. Tris(hydroxymethyl)-aminomethane (electrophoresis purity grade) and Coomassie Blue G250 staining solution were purchased from BioRad (Hercules, CA).

2.2.2 Cell Growth and Protein Extraction

Escherichia coli (ATCC 47076) cells were ordered from the American Type Culture Collection (ATCC). Cells were grown in Luria Broth (LB) (BBL, Becton Dickinson) at 37 °C with constant shaking and harvested at 72 h. The cells were then washed with sterile water, lyophilized and stored at -20 °C before extraction.

E. coli extracts were prepared by a solvent suspension method as described previously [3]. In brief, proteins were extracted by probe tip sonication using 0.1% TFA as extraction solvent. About 100 mg lyophilized *E. coli* sample was suspended in 5 mL 0.1% TFA solution in a 15 mL centrifuge tube. For sonication, a Branson Sonifier

450 (VWR Scientific, Bridgeport, NJ), with the duty cycle set to 75% and the output control set to 5 was used. The *E. coli* cell suspension was sonicated for 1 min with the tube on ice and then centrifuged at 13, 000 g. Supernatant (i.e., the clear solution above the cell debris) was then transferred into fresh 1.5 mL vials. The extracts were filtered using a Microcon-3 (Millipore) with 3, 000 Da molecular mass cut-off and then concentrated to ~0.5 mL by SpeedVac.

2.2.3 Fractionation of Proteins from the Soluble *E. coli* Whole Cell Lysates by RP-HPLC

Solvent delivery and separations were performed on an Agilent (Palo Alto, CA) HP1100 HPLC system. The solvents used for reversed-phase HPLC were water (A) and acetonitrile (B) with 0.1% (v/v) TFA in both phases. The gradients were 2-20% B in 10 min, 20-40% B in 30 min, 40-55% B in 5 min, 55-65% B in 10 min, and 65-90% B in 10 min. After the column was equilibrated for 30 min with 100% mobile phase A, 70 μL of bacterial extract was separated on a 250 × 4.6 mm i.d. C_8 column (5 μm particles with 300 Å pore size, Vydac) at a flow rate of 500 μL/min. 1 min fractions were collected with a Gilson FC 203B fraction collector (Gilson, Middleton, WI), and then concentrated to about 10 μL by SpeedVac.

2.2.4 In-solution Digestion

Each of the selected HPLC fractions was adjusted to pH 8.5 with 1M NH_4HCO_3 and then ~0.8 μg trypsin was added to each. Digestion was performed at 37 °C for ~2 h.

2.2.5 LC ESI MS/MS

Peptides in the in-solution protein digest of each HPLC fraction were desalted using commercial C18 μZiptips from Millipore (Bedford, MA) prior to LC ESI tandem mass spectrometry analysis. LC ESI MS/MS was performed on an LCQ Deca quadrupole ion trap mass spectrometer equipped with a dynamic nanospray source (ThermoFinnigan, San Jose, CA). The dynamic nanospray source was coupled to a Surveyor HPLC system (ThermoFinnigan, San Jose, CA). The flow from the pump was reduced from 300 μL/min to 1 μL/min by using a splitting tee and a length of restriction tubing made from fused silica. All separations were performed on a 150 μm (i.d.) ×15 cm capillary column, packed with 5 μm 218MS (C_{18}) beads (Vydac, Hesperia, CA). And 2 μL sample solution was injected for each run. The HPLC gradient was 5-10% B in 2 min, 10-45% B in 38 min, 45-65% B in 10 min, and followed by 65-85% B in 5 min (Solvent A, 0.5% acetic acid in water; B, 0.5% acetic acid in acetonitrile, v/v). The nanospray tip used was a 50 μm i.d. tip from New Objective (Woburn, MA). During the HPLC separation, the ion trap repetitively surveyed the full scan MS over the m/z range of 400-1800 and executed data-dependent MS/MS scans. MS/MS spectra were acquired using a relative collision energy of 30% (LCQ instrumental settings). An isolation width of 2 m/z units was used and recurring ions were dynamically excluded after two MS/MS spectra were obtained. Interpretation of the resulting MS/MS spectra was performed by the Sequest software. The *E. coli* proteome database created from a non-redundant protein database, which was downloaded from NCBI (http://www.ncbi.nlm.nih.gov), was used for database searching.

2.2.6 1D SDS PAGE

Proteins in each of the selected HPLC fractions were also separated using Tris-Tricine 10-20% gradient Ready Gels (Bio-Rad Laboratories, Hercules, CA). The premixed running buffer composition of the Tris/Tricine/SDS was as follows: 100 mM Tris, 100mM Tricine and 0.1% SDS. Sample buffer composed of 200mM Tris-HCl, 2% SDS, 40% Glycerol and 0.04% Coomassie G-250. Gel electrophoresis was performed at a constant voltage of 100 V. For some fractions, to increase the sample quantity, the same fractions from two separate runs were combined for 1D SDS-PAGE mini-gel separation. Proteins in the gel bands were visualized by staining with Coomassie-blue G-250.

2.2.7 In-gel digestion

Low mass gel bands of interest were excised and subjected to tryptic digestion according to the published digestion protocol [4]. Reduction was performed at 56 °C for at least 30 min in 10 mM DTT in 100 mM NH_4HCO_3. Alkylation was performed in the dark at room temperature for 30 min in 55 mM iodoacetamide in 100 mM NH_4HCO_3. Just sufficient 12.5 ng/μL trypsin in 50mM NH_4HCO_3 was added to cover the gel piece. Trypsin digestion was performed overnight (~12 h) at 30 °C. Sequential extraction of peptides from the gel was done twice with 0.25% TFA/20% ACN, twice with 0.25% TFA/50% ACN and once with 0.25% TFA/80% ACN. The extracts were pooled and the organic solvent removed by SpeedVac.

2.2.8 MALDI MS

MALDI MS results were obtained using a Bruker Reflex III MALDI time-of-flight (TOF) mass spectrometer (Bremen/Leipzig, Germany) equipped with a SCOUT 384 multiprobe inlet and a 337 nm nitrogen laser operated with a 3 ns pulse. Spectra were acquired in positive ion mode with delayed extraction using reflectron mode. The sample spot was scanned with the laser beam under video observation and spectra were acquired by averaging 300-500 individual laser shots and processed with the Bruker supporting software. The spectra were then reprocessed using the Igor Pro software package (WaveMetrics, Lake Oswego, Oregon, USA). Each spectrum was normalized to the most intense signal in the mass range displayed.

For LC/off-line MALDI analysis, each HPLC fraction was concentrated by a factor of 50 times to ~10 μL before mixing with matrix for analysis. A two-layer method was used for MALDI sample preparation [5]. HCCA was used as a matrix.

About 1 μL of 100 mM HCCA in acetone/methanol (130:45, v/v) was applied to the MALDI plate to quickly form the first layer. For the second layer, the sample solution was mixed at a volume ratio of 1:2 with saturated HCCA in formic acid/2-propanol/water (1:2:3, v/v/v). About 1 μL of the second layer solution was then applied onto the first layer and allowed to dry. On-spot washing of the MALDI sample with water was performed to remove any salts. Quasi-internal mass calibration was done as follows to improve the accuracy. The sample and standard were spotted close to each other on the same spot. The spectrum was collected from the sample spot first. After the intensity of the signal reached a certain level, the plate was moved and the laser beam was aimed at the standard. Finally, a spectrum, including peaks from both sample and standard, was recorded. Singly, doubly and triply charged peaks from cytochrome c were used to make this quasi-internal calibration. For MALDI MS on digests from in-gel tryptic digestions, the spectra were internally calibrated with trypsin autolysis peptide peaks and matrix peaks.

2.2.9 MALDI MS/MS

MALDI MS/MS spectra of the selected peptide peaks from the MALDI MS spectra of in-gel tryptic digests were acquired on an Applied Biosystems/MDS-Sciex QSTAR Pulsar QqTOF instrument equipped with an orthogonal MALDI source employing a 337 nm nitrogen laser (Concord, ON, Canada). The instrument was operated in positive ion mode and collisional induced dissociation (CID) of peptides was achieved with argon as the collision gas. Spectra were acquired and processed using Sciex supplied software and re-processed with Igor Pro software for presentation.

2.2.10 Possible Matches in Public Proteome Database

Possible matches in the public proteome database with the apparent masses observed in MALDI MS were searched as previously described [6]. The SWISS-PROT and TrEMBL databases were searched using the Sequence Retrieval System (SRS). The molecular mass measured by MALDI MS was entered into the information field, along with the organism *Escherichia coli*. Proteins were accepted as possible matches only if the database molecular masses matched with the experimental masses to within ±0.05%.

2.2.11 MALDI MS and MALDI MS/MS Mass Spectra Interpretation and Database Searching

Known contaminant peak masses were eliminated from each MALDI MS spectrum. The sample peaks were determined by comparing the peaks from a sample to the peaks from a blank piece of gel. Only sample peaks were considered for database searching. Artificial modifications of peptides by electrophoresis, such as, acrylamide adducts to cysteine, or oxidation of methionine were also considered for the database searching. Both peptide mass figureprinting (PMF) and the peptide sequencing results were searched for protein identification using the Mascot search program (http://www.matrixscience.com) and the UCSF Protein Prospector Database (http://prospector.ucsf.edu). The obtained partial sequence information for each peptide was used to confirm, discard or correct the previously obtained results from PMF.

2.3 Results and Discussion

In this study, an approach, shown schematically in Figure 2.1, for the characterization of PTMs of low-mass proteins observed in MALDI MS spectra was developed. The proteins in the cell extracts of *E. coli* were separated by RP-HPLC. The apparent molecular masses of low-mass proteins in each HPLC fraction were detected using MALDI MS. Then the proteins in each selected HPLC fraction were identified using two different tandem MS experiments. First, proteins in the selected fraction were digested in solution by trypsin. The resulting peptide mixture was separated by capillary column RP-HPLC using a gradient of acetic acid/acetonitrile. Individual peptide ions were automatically selected for CID tandem MS/MS followed by database searching using the SEQUEST software. Second, proteins in each selected HPLC fraction were separated by one-dimensional SDS polyacrylamide gel electrophoresis (1D SDS PAGE). Following electrophoretic separation, individual low-mass gel bands were digested *in-situ* by trypsin. Peptides were then extracted from the gel and each peptide mixture was analyzed by MALDI MS and MALDI MS/MS, followed by database searching using the MASCOT software. Finally, the identified proteins in each selected fraction were assigned to the apparent masses observed in the MALDI mass spectrum and any possible PTMs were assigned by comparing the

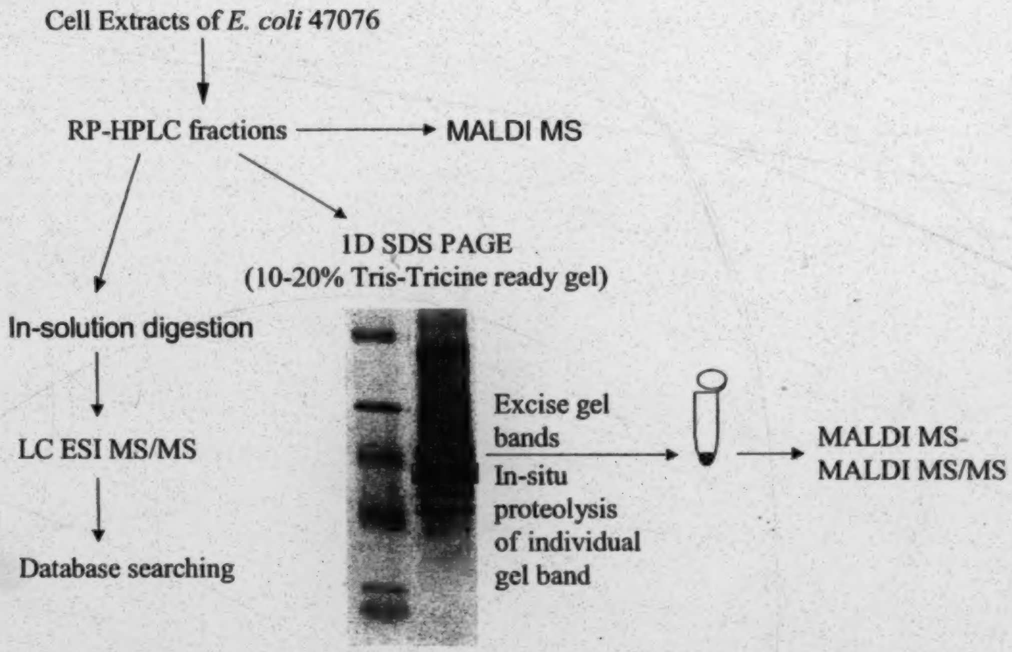
Cell Extracts of *E. coli* 47076
RP-HPLC fractions
MALDI MS
1D SDS PAGE
(10-20% Tris-Tricine ready gel)
In-solution digestion
LC ESI MS/MS
Database searching
Excise gel bands
In-situ proteolysis of individual gel band
MALDI MS
MALDI MS/MS

Figure 2.1 Schematic representation of the strategy used for annotating the origins of apparent masses observed in MS-derived *E. coli* 47076 protein mass database. Proteins identified by tandem MS/MS in each HPLC fraction were assigned to those apparent masses observed in the MALDI MS spectrum of the same fraction.

theoretical mass of the identified protein in the public database with the observed mass in MALDI MS analysis. Figure 2.2 shows the chromatogram of cell extracts of E. coli 47076. Since the major components eluted between 35 min and 52 min, efforts were concentrated on those fractions. The apparent molecular masses observed in MALDI MS spectra of HPLC fractions collected between 35 min and 52 min, which had signal to noise ratios over 5, are listed in Table 2.1. For proteins which appeared in several consecutive fractions, Table 2.1 only lists their initial fraction. The possible matched proteins in the SWISS-PROT and TrEMBL databases, searched using SRS, are listed in the third column in Table 2.1. Most of the apparent molecular masses observed in MALDI MS spectra have more than one match in the public proteome database. Accession numbers and names of proteins identified by combining the LC ESI MS/MS and MALDI MS/MS experiments, which correspond to those observed in MALDI mass spectra, are listed in the last two columns of Table 2.1. Each of these proteins was identified by the fragmentation patterns of at least two peptides. Table 2.1 shows that nine of the identified proteins have masses that are different from those in the public proteome database. This is because some of the proteins in the public proteome database are post-translationally modified *in vivo* or processed *in vitro* during the sample preparation process.

Fragmentation is one of the PTMs. Protein fragments have been identified by Edman sequencing of gel-separated *E. coli* proteins [7, 8]. The N-terminal sequence tags of these fragments matched the predicted internal region of the genes in the *E. coli*

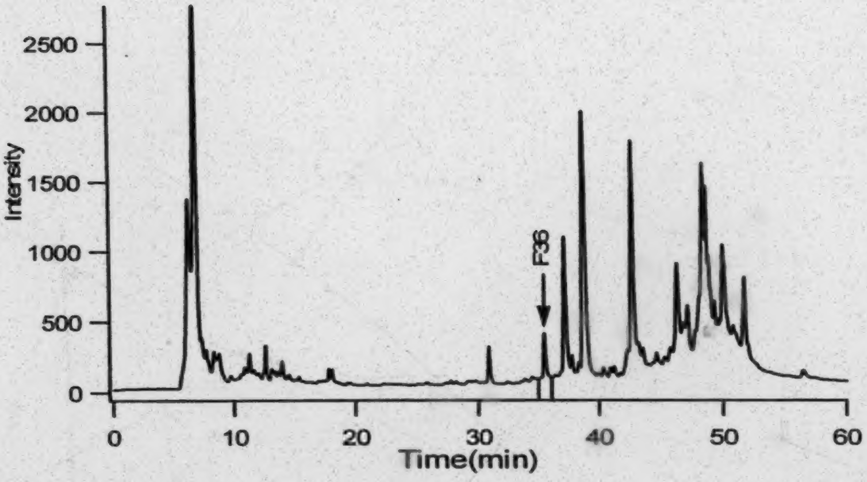

Figure 2.2 Chromatogram of cell extracts of *E.coli*. Note: F36 indicates the fraction collected between 36min and 37 min.

Table 2.1 Protein masses of protonated species observed in MALDI MS spectra of HPLC fractions of *E.coli* cell extracts, and tentative protein assignments based on the molecular mass and matched proteins and possible PTMs identified by combining LC MSMS and MALDI MSMS.

Fraction #	$M+H^+$ in MALDI MS	Acc Number of Possible Match	Mass of Possible Match	Identified Protein Acc Number	Identified Protein Name	PTM
35	8370	P05702	8372	P02379	30S ribosomal protein S21	
		P33649	8371			
		P02379	8369			
		Q8ZI69				
		P52038	8372			
		Q8X4G7	8366			
36	8876	P02428	8875	P02428	50S ribosomal protein L28	
		P76521	8874			
		Q8X9N0	8876			
		Q8X8R3	8876			
		Q8X720	8873			
36	8327	P05794	8328	P32691	Protein yjbJ	
		P02937	8323			
		P32691	8325			
		Q9L6J8	8325			
		Q8X675	8328			
		Q8X498	8327			
36	9080	Q9X5P1	9083	P26605	Protein hdeB	loss of signal peptide and oxidation
36	10386	P05380	10387	P02934	OMPA	fragment
		P97104	10388			
		Q8XBH5	10391			
36	11453	P02370	11449	P02370	30S ribosomal protein S14	
		Q9RPJ4	11455			
		Q46693	11454			
		Q8VV17	11453			
		Q8X9V3	11454			
37	7273	P15277	7272	P36996	Cold shock-like protein cspC	
		P37410				
		P36996	7271			
		O68636				
		P02429	7273			
		P58096	7273			

		Q46774	7275			
		P76136	7272			
		Q9EYC3	7268			
37	9064	Q8RTD2	9064	P26605	Protein hdeB	loss of signal peptide
		Q8XBG7	9059			
		Q8X4J7	9058			
		P76358	9065			
38	9756	P41039	9756	P26604	Protein hdeA	loss of signal peptide and oxidation
		Q47641	9753			
39	9584	P13479	9582			
		Q9R602	9580			
		P77387	9580			
		P77204				
39	9740	Q47037	9737	P26604	Protein hdeA	loss of signal peptide
		Q8XAC1	9742			
		Q8X4M7	9743			
		Q9R605	9734			
39	10300	P75877	10300	P02375	30S ribosomal protein S19	
		Q9R7Q1				
		P02375	10299			
		P32700	10303			
		Q8RTC2	10299			
		Q8XB54	10294			
		Q8X9M2	10297			
39	10695	P77618	10689	P02426	50S ribosomal protein L25	
		O54336				
		P02426	10693			
		Q8VQR3	10692			
		Q8VNS4	10693			
40	7334	P36997	7332	P36997	Cold shock-like protein cspE	
		P80434				
		P77103				
		P77098	7332			
		Q03535	7333			
		Q8X3H6	7334			
		Q8X378	7329			
40	8898	Q8XDS8	8899	P02374	30S ribosomal protein S18	acetylation
41	6242	P02436	6240			
		Q9AGQ2	6242			

		P77224	6241			
		Q8XCC4	6241			
		Q8X479	6238			
41	6412	P02430	6411	P02430	50S ribosomal protein L30	
43	5458	Q47654	5457			
		Q51953	5457			
43	6600	Q9RM53	6600			
		P37770	6599			
43	11978	Q8X7A6	11973	P52083	Protein ygiW	loss of signal peptide
		P71284	11982			
		Q8X6M6	11976			
		Q8X549	11980			
		Q8X3N1	11982			
44	7865	P33230	7863			
		P77683				
		P58033	7861			
44	12775	P04737	12768	P02419	50S ribosomal protein L18	
		P14517				
		P02419	12770			
		P22539	12780			
		P76367	12778			
		O07995				
		O07992				
		P31445	12779			
		O32571	12777			
		Q8X6T4	12771			
		Q8X4F0	12774			
		O88118	12771			
45	6699	P77370	6694			
		Q8X2R5	6701			
		Q8X2C4	6697			
45	9192	Q52277	9195	P02372	30S ribosomal protein S16	
		P02372	9190			
		P77006				
		P76575	9193			
		Q9AJD9	9190			
		P97155	9188			
		Q47583	9191			
		Q46853	9189			
45	11037	P25521	11035	P25521	Hfq protein	

		Q47383			
		P37668	11033		
		O82887	11030		
		Q8RTD5	11038		
		Q8X741	11039		
46	6509	P22986	6507		
		P77441			
		Q9EYC5	6504		
		Q8X3Y8	6507		
		Q9KXD1	6509		
46	9191	P02372	9190	P02372	30S ribosomal protein S16
		P77006			
		P76575	9193		
		Q9AJD9	9190		
		P97155	9188		
		Q8VR95	9185		
		Q47583	9191		
		Q46853	9189		
46	12010	P08338	12009		
		P08372	12004		
		P17577	12015		
		P09994			
		O52669	12012		
		Q9F573	12015		
		Q8VR73	12015		
		Q8X9R5	12004		
		Q8X894	12008		
46	12654	P77326	12658	P11285	Protein yfiA
		P11285	12653		
		Q9X730	12659		
		P71177	12655		
		P75729	12658		
47	9536	P02342	9535	P02342	DNA-binding protein HU-alpha
		Q9R601	9537		
		Q8X3G5	9533		
48	9385	P46132	9386		
		P77447			
		P22847	9381		
		Q93CH7	9388		
		Q9Z9I6	9383		
		Q8X8C1	9382		

48	10651	P08756	10651	P08756	Integration host factor beta-subunit	
49	11223	P07008	11225	P06984	Histone H1A, sperm	loss of N-terminal methionine
		Q8X3Y6	11226			
		Q8X5G9	11218			
		Q8X2D0	11224			
49	15327	P24233	15332	P02414	50S ribosomal protein L16	methylation and hydroxylation
		P03852	15322			
		O06959	15332			
		Q46764	15325			
		Q8XBH1	15324			
49	15409	P08936	15408	P08936	DNA-binding protein H-NS	
		P75684	15411			
50	9226	P02341	9225	P02341	DNA-binding protein HU-beta	
		P13966	9227			
		Q52770				
		Q8X8Q7	9226			
50	9271	P08365	9272	P32164	Hypothetical protein yiiU	loss of N-terminal MTM
		P76803				
		Q9JMS6	9271			

genome. The fragmentations were attributed to *in vitro* artifacts of sample preparation, *in vivo* events, or translation products initiated at the internal sites of the genes. None of these putative cleavage sites matches with any known *E. coli* protease recognition sequences, although little is known of the target specificity of *E.coli* proteases [9-11]. In this work, efforts toward the identification of protein fragments by tandem MS/MS will be discussed. Using a program called Paws that was downloaded from http://prowl.rockefeller.edu/, mass searching was attempted to correlate the identified proteins in Table 2.1 with the unidentified low mass species in the MALDI mass spectrum. By entering a peptide mass, Paws will search over an entire protein sequence to determine if any stretch of the sequence has a mass that matches with the input mass. If the observed mass in MALDI matched with a part of the sequence covered by the identified peptide sequences, the low mass species observed in MALDI was considered as a fragment of the specific protein. Mass information from one dimensional gel electrophoresis was used to achieve a conclusive identification.

In fraction 36, the observed molecular mass of 10386 Da ($M+H^+$) is a fragment of outer membrane protein A (OMPA, accession number P02934) instead of those matched by SRS. Figure 2.4(A) shows the MALDI spectrum of the tryptic digests of the indicated gel band as shown in Figure 2.3. Peptides matched with OMPA are labeled with an asterisk in Figure 2.4(B). OMPA was identified based on the fragmentation patterns of 4 peptides. Figure 2.4(C) shows one representative MS/MS spectrum. This protein has the following amino acid sequence:

MKKTAIAIAVALAGFATVAQAAPKDNTWYTGAKLGWSQYHDTGFINNNGPTHE
NQLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGSVENGAYKAQGVQLTAKL
GYPITDDLDIYTRLGGMVWRADTKSNVYGKNHDTGVSPVFAGGVEYAITPEIAT
RLEYQWTNNIGDAHTIGTRPDNGMLSLGVSYRFGQGEAAPVVAPAPAPAPEVQT
KHFTLKSDVLFNFNKATLKPEGQAALDQLYSQLSNLD*PKDGSVVVLGYTDR***IGSD**
AYNQGLSERRAQSVVDYLISKGIPADKISARGMGESNPVTGNT*C*DNVKQRAAL
ID*C*LAPDRRVEIEVK**GIKDVVTQPQA.** The bold characters show the sequence coverage of the MS/MS results. The molecular mass of this protein is 37292 Da, which was not found in the MALDI spectrum (Figure 2.4 (A)). Simple calculation shows that

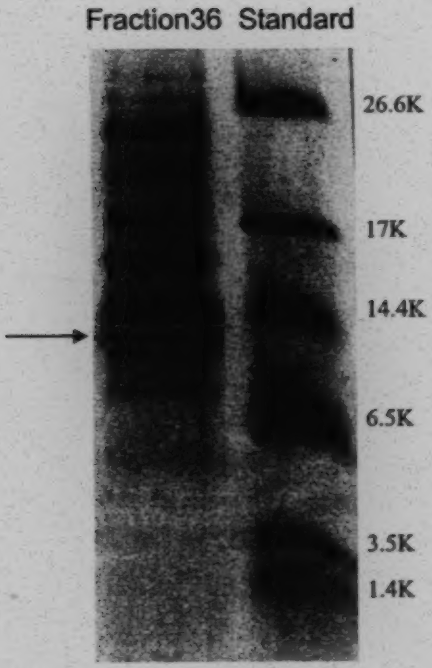
Fraction36
Standard
26.6K
17K
14.4K
6.5K
3.5K
1.4K

Figure 2.3 Proteome map of HPLC fraction 36 of *E. coli* cell extracts.

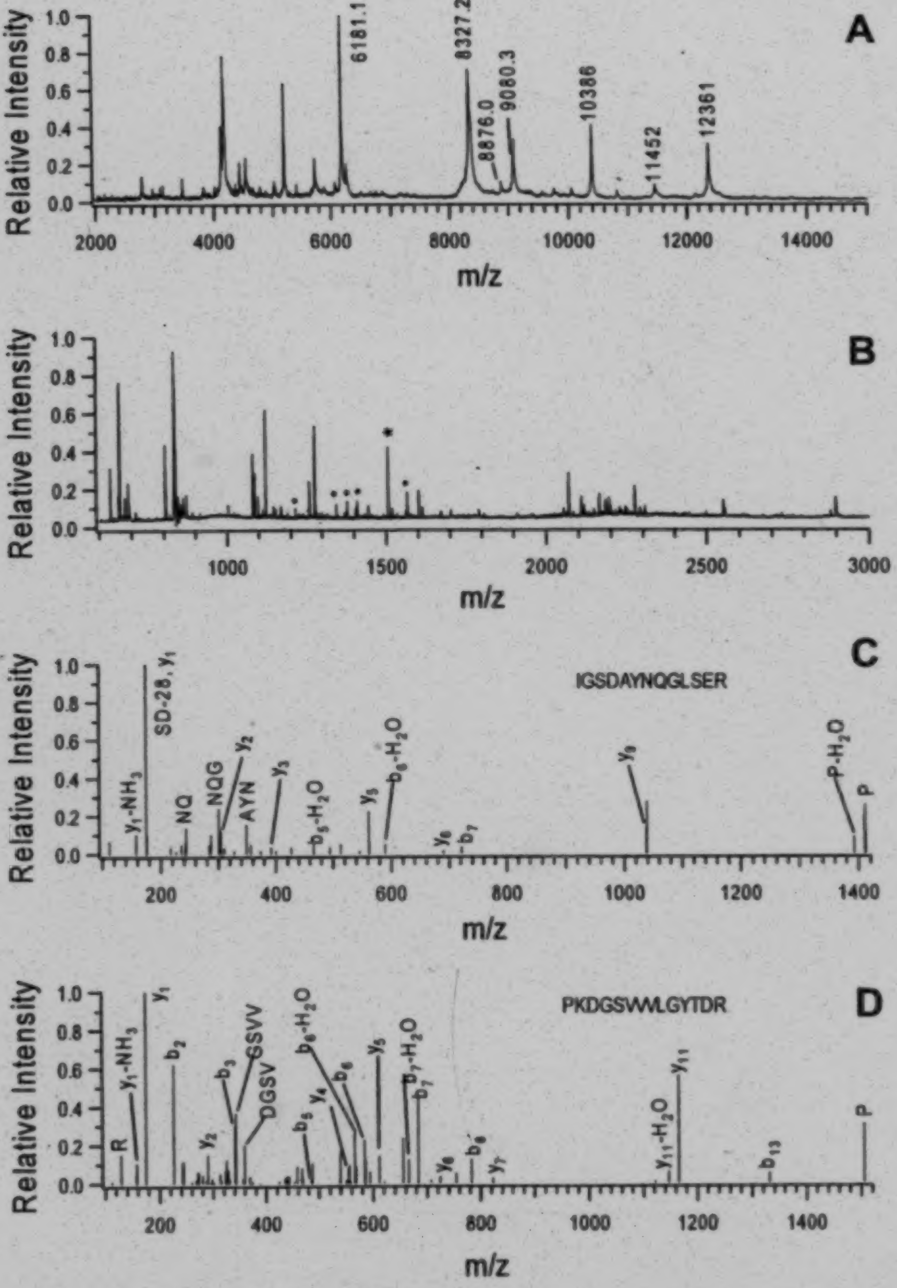

Figure 2.4 Fragmentation of OMPA observed in MALDI MS. (A) MALDI MS spectrum of HPLC fraction 36 of *E. coli* cell extracts. (B) Peptide map of tryptic digest of the band indicated in Figure 2.3. Peptides matched with that of OMPA are labeled with an asterisk (*). (C) LC MS/MS spectrum of peptide sequence of NNSLSQEVQNAQHQR. (D) LC MS/MS spectrum of peptide sequence of SLEVFEK.

the underlined protein sequence has the molecular mass of 10386.7 Da, matching a major component in Figure 2.4(A). This molecular mass information matches the molecular mass estimate of the gel. There is also a disulfide bond formed between two cysteines inside the underlined sequence. Therefore, the final molecular mass ($M+H^+$) of the underlined sequence is 10385.7 Da, which is very close to that observed in the MALDI mass spectrum. This was confirmed by the identification of an unexpected tryptic peptide PKDGSVVVLGYTDR (Figure 2.4(D)). Thus the apparent molecular mass 10386 Da corresponds to the fragment from OMPA. This was done by MASCOT searching without any enzyme type constraint. Seven y ions and seven b ions matched with the *in silico* fragmentation pattern, showing a strong correlation between the MS/MS spectrum and the identified peptide sequence. The fragmentation of this protein most likely occurred during protein extraction and isolation due to acid-labile cleavage of an amino acid bond between Asp (D) and Pro (P). It is found that in dilute acidic conditions, aspartyl peptide bonds tend to be more rapidly hydrolyzed than other aminoacyl peptide bonds, particularly for aspartyl proline bonds [12]. This is due to the neighboring group effects caused by the proximity of the side-chain carboxyl group of the amino acid residue Asp to the α-carboxyl peptide bond. On the basis of this notion, the fragmentation of this protein most likely occurred during protein extraction and isolation, since dilute TFA aqueous solution was used in both steps.

The most common PTM of the proteins in *E. coli* involves loss of the N-terminal methionine. For cytoplasmic proteins, the amino-terminal processing model [13, 14] predicts that the truncation of N-terminal Met residue by methionine aminopeptidase depends on the side-chain length of the second amino acid. This modification occurs most often in cases where the amino acid next to this methionine has a short side chain. Sterically large side chains are believed to prevent proteins from docking in the methionine aminopeptidase active site. When the second amino acid is Ala, Cys, Gly, Pro, Ser, Thr, or Val, the initiator Met is excised. Violations of the model were observed by Edman sequencing of the gel separated *E. coli* proteins [8] and it was believed that protein structures other than the second amino acid residue are involved in the excision specificity. It has been shown [8] that all initial Met residues are removed when the second residue is Ala or Ser and none of the Met is removed when Val is in the second position. The excision of Met is variable when the second residue is Thr, Gly, or Pro. It is very easy for the integration host factor alpha-subunit protein precursor (accession number P06984), observed in MALDI spectrum of fraction 49, to lose the first methionine in its sequence because the amino acid in the second position is alanine that has a small side chain.

The removal of signal peptides from protein precursors is another common PTM of the proteins in *E. coli*. Proteins located in the periplasm and outer membrane region are expected to have a signal sequence that helps direct their transport across the inner membrane [13, 15] The signal peptide generally has a positively charged amino-terminal region, a central hydrophobic region, and a carboxy-terminal region. *E. coli* signal peptides are 15 to 30 amino acids long and they are removed from the protein precursors (i.e., the gene expression products) by the signal peptidase, Lep, after transport through the membrane [13]. Therefore, the mature protein has a molecular mass 1500 to 3000 Da lower than predicted from the genome. The proteins with masses of 9740, 9064 and 11978 Da listed in Table 2.1 were identified to correspond to HDEA (accession number is P26604), HDEB (accession number is P26605) and YGIW (accession number is P52083) with the removal of their signal peptides. Additionally, the removal of the signal peptides from HDEA, HDEB and YGIW was confirmed previously by Link et al. [8] via N-terminal Edman degradation following separation of these proteins by 2D gel electrophoresis.

Loss of MTM (Methionine, Threonine, Methionine) from an amino terminus is a novel PTM observed in this study. In fraction 50, none of the proteins identified by LC MS/MS matched with the molecular mass 9271 Da observed in the MALDI mass spectrum (Figure 2.5(A)). One possible reason was that this molecular mass corresponded to a post-translationally modified protein. So SEQUEST database searching was done on LC-MS/MS spectra without an enzyme type constraint. Four y ions and four b ions matched with the *in silico* fragmentation pattern of the peptide SLEVSEK (Figure 2.5(C)). This peptide sequence is a partial sequence of hypothetical protein YIIU (accession number P32164). YIIU was identified based on the fragmentation patterns of 3 peptides. Figure 2.5(B) shows one representative MS/MS spectrum. YIIU_ECOLI has the following sequence: MTMSLEVFEKLEAKVQQAIDTITLLQMEIEELKEKNNSLSQEVQNAQHQREE

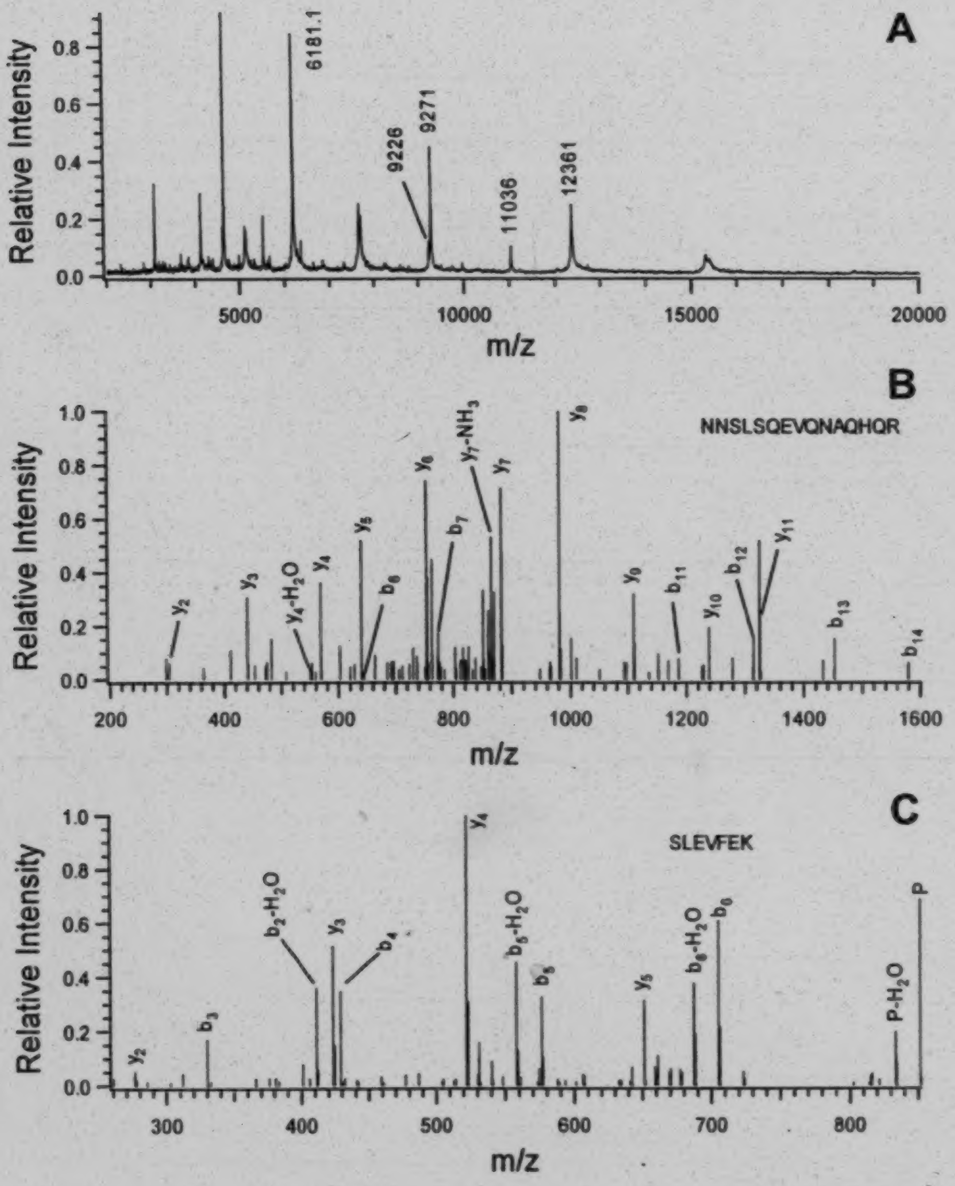

Figure 2.5 Hydroxylation of arginine in 50S ribosomal protein L16 observed in MALDI MS. (A) MALDI MS spectrum of HPLC fraction 50 of *E. coli* cell extracts. (B) LC MS/MS spectrum of peptide with sequence of NNSLSQEVQNAQHQR. (C) LC MS/MS spectrum of peptide with sequence of SLEVFEK.

LERENNHLKEQQNGWQER**LQALLGRMEEV**. The bold sequence represents coverage by the MS/MS results. Using the Paws Program, it was found that the underlined protein sequence has an average molecular mass of 9271 Da, matching with one major component in the MALDI MS spectrum of HPLC fraction 50. This means that the apparent mass of 9271 Da observed in MALDI MS is YIIU_ECOLI with the loss of MTM at the N-terminus.

Several other PTMs, including methylation, oxidation, and acetylation are also listed in Table 2.1. In fraction 49, none of the proteins identified by LC MS/MS and MALDI MS/MS matched with the molecular mass 15326 Da observed in the MALDI mass spectrum (Figure 2.6(A)). It was reported that 50S ribosomal protein L16, which has a nominal molecular mass of 15281 Da based on the genome sequence, was seen to have an apparent mass of 15326 Da, 45 Da higher, in its MALDI mass spectrum [16]. 50S ribosomal protein L16 was identified based on the fragmentation patterns of 4 peptides. Figure 2.6(B) shows one representative MS/MS spectrum. The N-terminal methionine of this protein has been reported as being methylated [17]. The original sequencing of this subunit also indicated that the arginine residue at position 81 was modified in some way. Subsequent work by J. Brosius suggested it was hydroxylated [16]. This is confirmed by MALDI MS/MS on the selected parent peptide ion of 1726.01 Da (VFPDKPITEKPLAVR, Figure 2.6(C)), which correlates to 1710.00 Da and the addition of 16 Da, due to hydroxylation of an arginine residue at position 81 of the sequence of 50S ribosomal protein L16. When the MS/MS spectrum of the parent ion

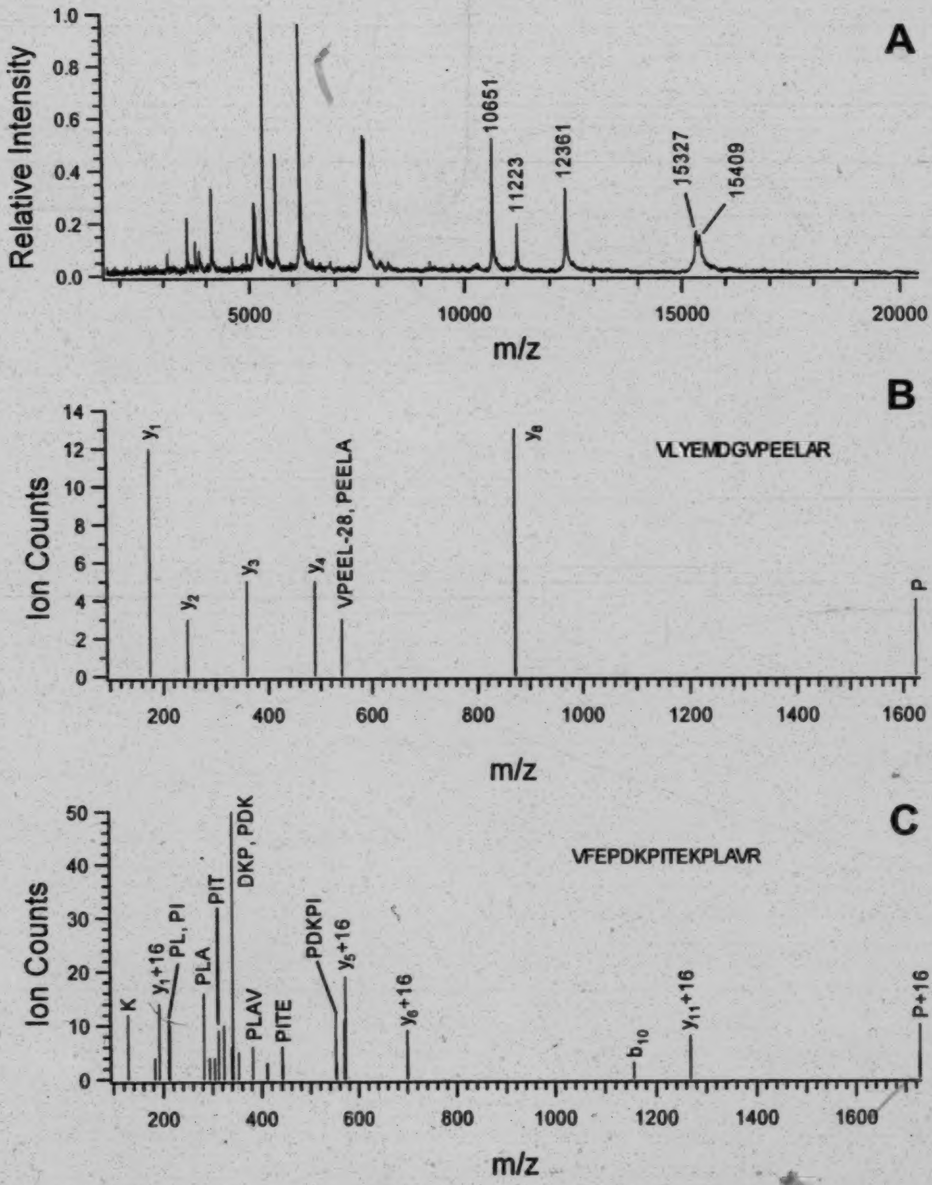

Figure 2.6 Loss of MTM in hypothetical protein YIIU_ECOLI observed in MALDI MS. (A) MALDI MS spectrum of HPLC fraction 49 of *E. coli* cell extracts. (B) MALDI MS/MS spectrum of peptide sequence of VLYEMDGVPEELAR. (C) MALDI MS/MS spectrum of peptide sequence of VFEPDKPITEKPLAVR.

of 1726.0 Da was searched on MASCOT, there were no significant matches. After 16 Da was added to arginine, the spectrum matched with the product ion of peptide VFPDKPITEKPLAVR. These two modifications still yield a protein that is 15 Da lighter than the observed mass, suggesting that another methylation or hydroxylation is present. Efforts to confirm this were unsuccessful because another peptide that has molecular mass 16 or 14 Da higher than the theoretical mass could not be found.

2.4 Conclusions

Using the described approach, not only previously reported PTMs involving acetylation, methylation, oxidation and the removal of signal peptides, but also several novel PTMs, such as loss of N-terminal Met-Thr-Met (MTM) and hydroxylation of arginine, were observed. In this study, 37 low mass proteins were observed in the MALDI mass spectra of HPLC fractions of *E. coli* cell extracts and 28of them were identified. Among the identified proteins, 9 were modified, resulting in mass shifts between the theoretical masses and the apparent masses in MALDI MS. While the current method was developed and applied to the analysis of PTMs of proteins extracted from a relatively well-characterized microorganism, it is envisaged that this method should be applicable to other biological systems for the discovery of new PTMs.

2.5 Literature Cited

(1) Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255-261.

(2) Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198-207.

(3) Wang, Z.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 456-464.

(4) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850-858.

(5) Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087-1091.

(6) Dunlop, K. Y.; Li, L. *J. Chromatogr. A* **2001**, *925*, 123-132.

(7) Wasinger, V. C.; Humphery-Smith, I. *FEMS Microbiol Lett* **1998**, *169*, 375-382.

(8) Link, A. J.; Robison, K.; Church, G. M. *Electrophoresis* **1997**, *18*, 1259-1313.

(9) Maurizi, M. R. *Experientia* **1992**, *48*, 178-201.

(10) Lazdunski, A. M. *FEMS Microbiol. Rev.*. **1989**, *5*, 265-276.

(11) Miller, C. G.; Neidhardt, F. C.; Curtiss, R.; Gross, C. M.; Schaechter, M.; Umbarger, H. E. E., Eds. *Escherichia coli and Salmonella: Cellular and Molecular Biology*; ASM Press: Washington, DC, 1996.

(12) Piszkiewicz, D.; Landon, M.; Smith, E. L. *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1173-1178.

(13) Murphy, C. K.; Beckwith, J.; Neidhardt, F. C.; Curtiss, R.; Gross, C.; Ingraham, J. L.; Lin, E. C.; Low, K. B.; Magasanik, B.; Riley, M.; Schaechter, M.; Umbarger, H. E. **1996**, 967-978.

(14) Ben-Bassat, A.; Bauer, K.; Chang, S. Y.; Myambo, K.; Boosman, A.; Chang, S. *J. Bacteriol.* **1987**, *169*, 751-757.

(15) Hirel, P. H.; Schmitter, M. J.; Dessen, P.; Fayat, G.; Blanquet, S. *Proc. Natl. Acad.*

Sci. U. S. A. **1989,** *86*, 8247-8251.

(16) Arnold, R. J.; Reilly, J. P. *Anal. Biochem.* **1999**, *269*, 105-112.

(17) Brosius, J.; Chen, R. *FEBS Lett.* **1976,** *68*, 105-109.

Chapter 3

Quantitative Proteome Analysis Using Differential Stable Isotopic Labeling and Microbore LC-MALDI MS and MS/MS [a]

3.1 Introduction

Detection and quantification of differences in the protein profiles of cells, tissues or body fluids of different origins or states is increasingly being recognized as a key objective of proteomics research [1]. The measurement of differential protein expression, resulting from physiological or environmental perturbation, provides a more direct, accurate, and versatile way to detect global changes in cellular dynamics in health and disease, compared to the complementary and more mature technology of mRNA expression analysis [2]. Quantitative proteomics plays a pivotal role in the discovery of diagnostic or prognostic protein markers, for the detection of new therapeutic targets, and as a powerful tool to further our understanding of basic biological processes and mechanisms.

Traditionally protein quantification is done by using gel-based methods. In the past few years, several methods based on stable isotopic labeling of proteins or peptides have been reported for determining the relative abundance of proteins expressed between two different states of a similar biological system [3-7]. Among these methods, the isotope-coded affinity tag (ICAT) approach pioneered by Aebersold and co-workers [3, 8-12] has been extensively used. The main advantage of this method is that it enriches peptides containing the rare amino acid cysteine, thereby significantly reducing the complexity of the peptide mixture and increasing the dynamic range of MS analysis [13, 14]. On the other hand, the use of the ICAT reagents fails for quantification of cysteine-free proteins and the cost is another negative aspect of the ICAT reagents. Therefore, other alternative labeling reagents for peptides or metabolically labeling of proteins have been recently developed by several groups for the quantification of global protein expression [4, 5, 7, 15-30].

---

[a] A portion of this chapter is published as: C. Ji, and L. Li, "Quantitative Proteome Analysis Using Differential Stable Isotopic Labeling and Microbore LC-MALDI MS and MS/MS" *J.Proteome Res.*, **2005**, *4*, 734-742.

To date, most developed quantification strategies based on stable isotope coding have been carried out using reversed-phase microcapillary liquid chromatography (RP-μLC) coupled online with electrospray ionization (ESI) MS and MS/MS, in which peptides at specific mass-to-charge (*m/z*) values are selected for collision-induced dissociation (CID) using data-dependent software routines and are identified by searching the resulting CID spectra against a protein sequence database. The abundance ratio is determined for each identified protein. This procedure is robust, sensitive, and can be automated. However, it also has some drawbacks. The analysis of complex, isotope-tagged peptide mixtures by on-line RP-μLC ESI MS/MS suffers from the demand for continual sample consumption and the untargeted selection of precursor ions for sequencing. Therefore, during the analysis of complex protein mixtures, potentially large numbers of proteins are identified that do not show a quantitative change under the conditions tested and may thus be of limited interest. In addition, multiple charged ions generated by ESI complicate the MS spectra and require a large mass difference between isotopically coded pairs to avoid peak overlaps. This is particularly true for low resolution MS and MS/MS instruments such as ion trap MS.

Recently, Griffin and co-workers developed an approach for the quantification and identification of the components of complex protein mixtures based on µLC-matrix-assisted laser desorption/ionization (MALDI) quadrupole-time-of-flight (QqTOF) MS in conjunction with the ICAT technology [31, 32]. The use of MALDI MS and MS/MS offers several advantages. MALDI spectra are mainly composed of singly charged peptide peaks, which simplifies peak picking and quantification. Temporal constraints of on-line detection that are encountered with LC-ESI MS are eliminated as the peptides separated by µLC are deposited on a MALDI sample plate prior to MS analysis. Additionally, targeted protein identification can be carried out by MALDI MS/MS on peptides that show abundance changes in MALDI MS analysis, resulting in high efficiency in sample utilization and MS instrument time.

More recently, Zappacosta and Annan have reported an alternative labeling strategy to ICAT that has been demonstrated to be useful for protein quantification by µLC MALDI-TOF MS [33]. Protein identification was done by both LC-ESI MS/MS and LC-MALDI TOF/TOF. In their approach, N-terminus of all peptides were labeled with either a d(5) or d(0)-propionyl group using d(10)-propionic or d(0)-propionic anhydride after lysine side-chain amino groups were blocked by guanidination. Thus the technique should be suitable for global proteome quantification.

Although LC combined with either ESI MS or MALDI MS is being used for quantitative proteome analysis, its sample loading capacity is limited. The use of a larger column, hence a greater sample loading, is expected to increase the ability of detecting low abundance proteins in a complex mixture [31]. For protein quantification, this becomes even more important. To reliably determine the abundance ratio of peptide pairs, the peak signal intensities should be much greater than the background signals. For example, the detection limit for judging the presence of an analyte signal is commonly defined as analyte-signal-to-background ratio of greater than 3. However, the limit of quantification is usually defined as signal-to-background ratio of greater than 10. In addition, a larger sample loading offers a greater chance of detecting multiple peptides from a protein, thereby resulting in better statistics for protein quantification based on the average value of relative abundances of more than one pair of peptides. Thus a combination of a large column separation with MS should be particularly useful for quantitative proteomics.

Li and his coworkers have recently developed a heated droplet MALDI interface which can work with high flow LC separations [34]. Since MALDI signals are dependent on the analyte concentration in the matrix crystals, not the analyte solution concentration, detection sensitivity of MALDI is determined by the total amount of analyte deposited per deposition area. Thus, a large column can be used in LC-MALDI so long as the elutes from LC are concentrated to small spots during LC deposition. With the heated droplet interface, solvents are effectively evaporated during sample deposition and the detection sensitivity of peptides in low fmol regime can be readily achieved. In Li's previous reports [34-36], he demonstrated that a greater number of peptides can be detected and sequenced by MALDI MS and MS/MS using microbore LC separation, compared to capillary LC. However, this heated droplet interface uses heat (~100 °C) to evaporate solvents from the hanging droplets at the exit of the capillary tube in the interface as well as the droplets landed onto the MALDI plate. Although thermal degradation of peptides and proteins including many phosphorylated peptides and glycopeptides, which have been analyzed, are not observed, thermal degradation can potentially take place for labeled peptides in quantitative proteomics. In this case, the labeling groups may dissociate from the peptides via thermal decomposition or thermally assisted chemical reactions such as hydrolysis.

In this study, the compatibility of two global labeling methods (i.e., esterification to carboxylic groups [16, 37] and dimethylation to amine groups of peptides [17, 38, 39]) for microbore LC-MALDI with the heated droplet interface for quantitative proteomics is evaluated. It is found that, using reductive amination with either d(0)- or d(2)-formaldehyde to label amine groups in peptides, protein relative abundances can be determined very reliably based on this LC-MALDI technique.

3.2 Experimental

3.2.1 Chemicals and Reagents

Anhydrous d(0)-methyl alcohol and d(3)-methyl alcohol, d(0)-formaldehyde (37 wt. % solution in H_2O), sodium cyanoborohydride, horse cytochrome *c*, human ubiquitin, horse myoglobin, bovine carbonic anhydrase II, bovine beta casein, bovine trypsin and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (Oakville, ON, Canada). d(2)-Formaldehyde (~20% w/w solution in deuterated water) was obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). Acetonitrile was purchased from Fisher Scientific Canada (Edmonton, AB, Canada). Water used in these experiments was obtained from a Milli-Q Plus purification system (Millipore, Bedford, MA). The MALDI matrix 2,5-dihydroxybenzoic acid (DHB) was purchased from Aldrich (Milwaukee, WI).

3.2.2 Preparation of Control Protein Mixture Digests

Two test control protein mixtures were prepared with each containing the same five standard proteins at different concentrations. The names of these proteins along with their abbreviated names as given in the Swiss-Prot annotated protein sequence database (http://ca.expasy.org/sprot/) are: human ubiquitin (UBIQ HUMAN), horse cytochrome *c* (CYC HORSE), horse myoglobin (MYG HORSE), bovine carbonic anhydrase II (CAH2 BOVIN) and bovine beta casein precursor (CASB BOVIN). The protein mixture A contains ubiquitin (3 nmol/mL), cytochrome *c* (3 nmol/mL), myoglobin (3 nmol/mL), carbonic anhydrase II (3 nmol/mL) and beta casein (3 nmol/mL). The protein mixture B contains ubiquitin (3 nmol/mL), cytochrome *c* (6 nmol/mL), myoglobin (12 nmol/mL), carbonic anhydrase II (1.5 nmol/mL) and beta casein (0.75 nmol/mL). Protein mixtures A and B were digested with trypsin overnight at 37 °C at an enzyme/substrate ratio of 1:40 (w/w). The digestion was stopped by adjusting pH to about 5 with 1% TFA.

3.2.3 *E. coli* Protein Extraction and Digestion

Bacterial extracts were prepared using TRIzol Reagent (Invitrogen Canada Inc.) according to the manufacturer instructions. The lyophilized bacterial cells (5 mg) were suspended in 0.5 mL of TRIzol Reagent in a 1.5 mL vial. The solution was pipetted up and down several times and incubated with the homogenized cells for 5 min at room temperature. Following cell lysis, 0.1 mL chloroform was added into the vial. Then the vial was shaken vigorously for 15 s and incubated at room temperature for 3 min. The vial was centrifuged at 12 000 × *g* for 10 min at 4 °C. The aqueous phase was dumped and DNA in the inter-phase and organic phase was precipitated with 0.15 mL 100% ethanol. Finally, proteins in the phenol-ethanol supernatant were precipitated with the addition of 0.75 mL of 2-propanol and the resulting protein pellet was washed 3 times with a solution containing 0.3 M guanidine hydrochloride (HCl) in 95% ethanol. The final protein pellet was dried using a speedVac. 1 mg *E. coli* whole cell extracts were resuspended in 100 μL of 6 M guanidine-HCl and diluted to 0.6 M guanidine-HCl. Finally, a small amount of freshly prepared trypsin solution (trypsin content equal to 1/100th of the total protein in the mixture) were added to the protein mixture, incubated at 37 °C for overnight, and then stopped by acidifying the solution. The resultant peptide mixture was desalted using 2.1-mm C_{18} guard column and kept in a -78 °C freezer until ready to use.

3.2.4 Preparation of Methyl Esters

Control protein mixtures of ubiquitin, cytochrome *c*, myoglobin, carbonic anhydrase II and β-casein were proteolyzed to peptides using trypsin. Prior to methylation, peptide solutions were lyophilized to dryness in a Speedvac. Lyophilized peptides were methylated after solubilization in a solution of methanolic HCl [37]. Methanolic 2 N HCl was prepared fresh daily by adding 160 μL acetyl chloride (Aldrich) to 1 mL of anhydrous d(0)-methyl alcohol or d(3)-methyl alcohol dropwise with stirring. After 5 min, 100 μL of the reagent was added to the lyophilized peptide mixtures. Esterification proceeded for 2 h at room temperature and the reaction was stopped by lyophilization to dryness. To obtain full conversion of carboxyl groups to methyl esters, the procedure was repeated. Methylated peptides were solubilized in 0.1% TFA for microbore LC-MALDI QqTOF analysis.

3.2.5 Reductive Amination

Control protein mixtures of ubiquitin, cytochrome *c*, myoglobin, carbonic anhydrase II and beta casein were proteolyzed to peptides using trypsin. The digestion was stopped by adjusting pH to about 5 with 1% TFA. Then equal volume of sodium acetate buffer (0.2 M, pH 6.0) was added into each vial containing control protein mixture digest, vortexed and mixed with freshly prepared sodium cyanoborohydride (1 M, 10 µL). The mixtures were vortexed again and mixed with either d(0)- or d(2)-formaldehyde (4% in water, 5 µL). The mixtures were vortexed and incubated at 37 °C for 1 h. If necessary, ammonium bicarbonate (1 M, 5 µL) was added to consume the excess formaldehyde.

3.2.6 Protein Mixtures for Concentration Dynamic Range Studies

Six samples with each containing different relative concentrations of ubiquitin and myoglobin were used to examine the capability of the microbore LC-MALDI approach for detecting low abundance proteins in the presence of high abundance proteins. The concentration of ubiquitin was the same in all mixtures, i.e., 2 µg/µL. The molar ratios between ubiquitin and myoglobin of the six protein mixtures (M1, M2, M3, M4, M5, and M6) were 1000:1 for M1, 500:1 for M2, 4000:1 for M3, 2000:1 for M4, 20 000:1 for M5 and 10 000:1 for M6. After digestion with trypsin, peptides in M1, M3, and M5 were labeled with d(0)-formaldehyde and peptides in M2, M4, and M6 were labeled with d(2)-formaldehyde. Then equal volumes of M1 and M2 were combined to produce a mixture for microbore-LC MALDI MS and MS/MS. In the same manner, M3 and M4 were combined and M5 and M6 were mixed for MS analysis.

3.2.7 Microbore LC-MALDI QqTOF Mass Spectrometric Analysis

The separated d(0)- and d(3)-methylated peptide mixtures or d(0)- and d(2)-dimethylated peptide mixtures were combined and the mixture separation was performed on an Agilent (Palo Alto, CA) 1100 series capillary HPLC equipped with an auto sampler. Chromatographic analysis was performed with a reversed-phase 1.0 × 150 mm Vydac C_{18} column (5 μm particles with 300 Å pore size, Catalog No.: 218TP5115). A flow-rate of 40 μL/min was used for separation. Gradient elution was performed with solvent A (Milli-Q water, 0.1% TFA and 4% acetonitrile in water, v/v/v) and B (0.1% TFA in acetonitrile, v/v). For the separation of the digests from the protein mixtures, the gradient was 0-10% B in 10 min, 10-40% B in 40 min, 40-90% B in 15 min. About 40 pmol (50 μL) labeled digests were injected. For the separation of the digests from the *E. coli* cell extract, the gradient was 0-10% B in 5 min, 10-40% B in 80 min, 40-90% in 10 min. HPLC fractions were directly collected in 1-min time intervals onto a 100-well MALDI plate (Applied Biosystems, Concord, ON, Canada) using a home-built, heated droplet LC-MALDI interface [34]. After the fractionation was completed, the dried peptides in each well was redissolved and mixed with DHB matrix by the addition of 0.8 μL of 100 mg/mL DHB matrix in 50%ACN/50%water (v/v).

MALDI MS and MALDI MS/MS data were acquired on an Applied Biosystems/MDS-Sciex QSTAR Pulsar QqTOF instrument equipped with an orthogonal MALDI source employing a 337 nm nitrogen laser (Concord, ON, Canada) that has been previously described [40]. The instrument was operated in positive ion mode and CID of peptides was achieved with argon as collision gas. Spectra were acquired and processed using Sciex supporting software and re-processed with Igor Pro software (WaveMetrics, Lake Oswego, Oregon, USA) for presentation.

3.2.8 Protein Identification from MS/MS Data

Peptide sequences were automatically identified by database searching of the MS/MS spectra against the Swiss-Prot database using the Mascot search program (http://www.matrixscience.com). In all cases, MS/MS data were searched twice in one case with no requirement that the peptides be tryptic and in the other case being constrained to only tryptic peptides. In both cases, the mass tolerance of the precursor peptide and its fragments was set at ±0.3 Da. All of the MS/MS spectra were manually checked to verify the validity of the Mascot results (see Results and Discussion).

3.2.9 LC-ESI MS

Isotope effect analysis was carried out in a Bruker/Agilent Esquire-LC Ion Trap LC/MS^n system. Myoglobin was digested to peptides with trypsin and the solution divided into four equal aliquots, two of which were esterified separately using d(0)- or d(3)-methanol and another two of which were dimethylated separately using d(0)- or d(2)-formaldehyde. After labeling, the d(0)- and d(3)-methylated myoglobin digests or d(0)- and d(2)-formaldehyde labeled myoglobin digests were combined and the mixtures were separated by RP-HPLC. The HPLC pump was operated at a flow rate of 100 μL/min and split to obtain flow through a 150 × 1 mm I.D. C_{18} column (Vydac) at 40 μL/min. As peptides eluted from the microbore column, they were electrosprayed directly into the ion trap mass spectrometer. Solvent delivery and separations were performed on an Agilent (Palo Alto, CA) HP1100 HPLC system. Gradient elution was performed with solvent A (Mili-Q water, 0.1% TFA and 4% acetonitrile, v/v/v) and B (0.1% TFA in acetonitrile, v/v). The gradient was 0-10% B in 10 min, 10-40% B in 40 min, 40-90% B in 15 min. Mass spectra were acquired over the mass range *m/z* 400-1800. All data were reprocessed using Bruker Daltonics DataAnalysis software.

3.3 Results and Discussion

Figure 3.1 shows the workflow for protein quantification and identification using global differential stable isotopic labeling and the microbore LC-MALDI QqTOF mass spectrometer. Two control protein mixtures (A and B) are initially tryptically digested, labeled with either the light or heavy forms of methanol for esterification or formaldehyde for dimethylation. The light and heavy labeled peptides are combined. The peptide mixture is then separated by microbore RP-LC column, with spotting of the

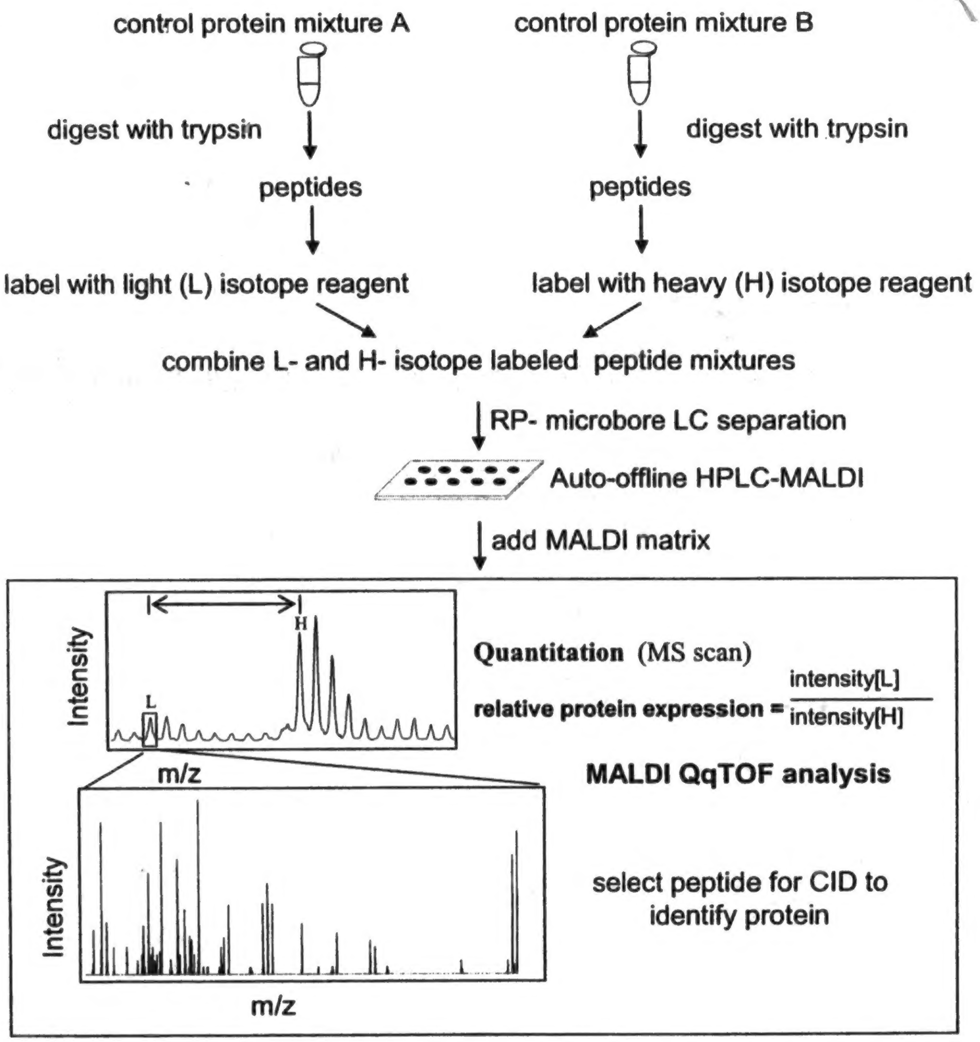
control protein mixture A
control protein mixture B
digest with trypsin
digest with trypsin
peptides
peptides
label with light (L) isotope reagent
label with heavy (H) isotope reagent
combine L- and H- isotope labeled peptide mixtures
RP- microbore LC separation
Auto-offline HPLC-MALDI
add MALDI matrix
Intensity
H
L
m/z
Quantitation (MS scan)
relative protein expression = intensity[L] / intensity[H]
MALDI QqTOF analysis
Intensity
m/z
select peptide for CID to identify protein

Figure 3.1 Workflow for quantitative protein analysis by combining global differential stable isotope labeling with microbore LC-MALDI QqTOF mass spectrometry.

eluent in discrete fractions onto a MALDI target using the heated droplet LC-MALDI interface. After LC fractionation and deposition, a MALDI matrix (DHB) is added to each spot on the target. This is followed by MALDI QqTOF mass spectrometric analysis of each sample spot. An initial MS scan is performed to quantify the global differential stable isotopic labeled peptide pairs by comparison of the monoisotopic signal intensities of the light and heavy labeled forms of the peptide. MALDI MS/MS analysis and sequence database searching are subsequently done on the peptide pairs to identify the proteins present.

From the workflow shown in Figure 3.1, one can see that, to develop a reliable quantitative method, several issues need to be addressed. These include the compatibility of labeling chemistry with the LC-MALDI interface, isotope effects of differentially labeled peptides on separation, precision on determination of the relative abundance of peptide pairs, and the effect of labeling chemistry on protein identification.

3.3.1 Comparison of Labeling Methods

To first determine the ability of microbore LC-MALDI MS to quantify and identify differential stable isotopic esterification or dimethylation labeled peptides, control mixtures of proteins of known identity and composition were analyzed. Two mixtures containing different molar concentrations of five different proteins were prepared as described in Section 3.2.6. These two mixtures were compared and analyzed using the experimental protocol illustrated in Figure 3.1. Tables 3.1 and

Table 3.1 Esterification labeling and microbore LC-MALDI QqTOF analysis results of two control protein mixtures.

Protein	# of peptide pairs identified	Abundance ratio [d(0)/d(3)]*			
		Ave	%SD	Exp'd	%Err
Myoglobin	13	3.99	9.8	4.00	0.25
Cytochrome c	8	2.07	14	2.00	3.5
Ubiquitin	4	1.02	11	1.00	2.0
Carbonic anhydrase II	5	0.49	8.2	0.50	2.0
Beta casein	5	0.26	12	0.25	4.0

*Ave: average abundance ratio calculated from all the peptide pairs; %SD: relative standard deviation; Exp'd: expected abundance ratio; %Err: percent error calculated from the observed average abundance ratio and the expected ratio.

Table 3.2 Dimethylation labeling and microbore LC-MALDI QqTOF analysis results of two control protein mixtures.

Protein	# of peptide pairs identified	Abundance ratio [d(0)/d(2)]*			
		Ave	%SD	Exp'd	%Err
Myoglobin	12	0.25	6.4	0.25	0
Cytochrome c	11	0.49	8.6	0.50	2.0
Ubiquitin	6	1.01	9.1	1.00	1.0
Carbonic anhydrase II	10	1.96	7.3	2.00	2.0
Beta casein	7	3.9	5.7	4.00	2.5

*Ave: average abundance ratio calculated from all the peptide pairs; %SD: relative standard deviation; Exp'd: expected abundance ratio; %Err: percent error calculated from the observed average abundance ratio and the expected ratio.

3.2 provide the summaries of the results obtained using esterification and dimethylation labeling, along with the number of peptide pairs that were identified by MS/MS database searching. The average relative intensity ratios of the d(0)- and d(3)-methanol labeled peptide pairs [d(0)/d(3)] or d(0)- and d(2)-formaldehyde labeled peptide pairs [d(0)/d(2)] determined by the initial MS scan along with the expected d(0)/d(3) or d(0)/d(2) values in the sample are listed. A representative result is shown in Figure 3.2. A segment of the MS scan of a specific sample spot is shown in Figure 3.2A with a differential esterification labeled peptide pair, where the d(0)- and d(3)-methanol labeled peptides have observed monoisotopic *m/z* values of 1544.642 and 1553.695 for their single-charged molecular ions, respectively. The d(0)-methanol labeled peptide was selected for CID, and the resulting MS/MS spectrum is shown in Figure 3.2B. Database searching of this spectrum using Mascot software matched this peptide to the sequence HPG**D**FGA**D**AQGAMT**K** from the horse myoglobin, where three carboxylic acid groups in the side chains of 2 D's and carboxyl terminal amino acid have been modified with d(0)-methanol. The mass difference (9.053 Da) observed in MS scan corresponds to the mass difference (9.054 Da) resulting from three esterification sites present in this peptide.

Although this sample was digested with trypsin, for unknown reasons, the proteins in the mixture were also cleaved frequently at nonspecific cleavage sites. This has also been observed by others [31]. Several unexpected peptides resulting from the nonspecific cleavage are included in the data shown in Tables 3.1 and 3.2, which were identified by database searching using Mascot without any constraint of enzyme type.

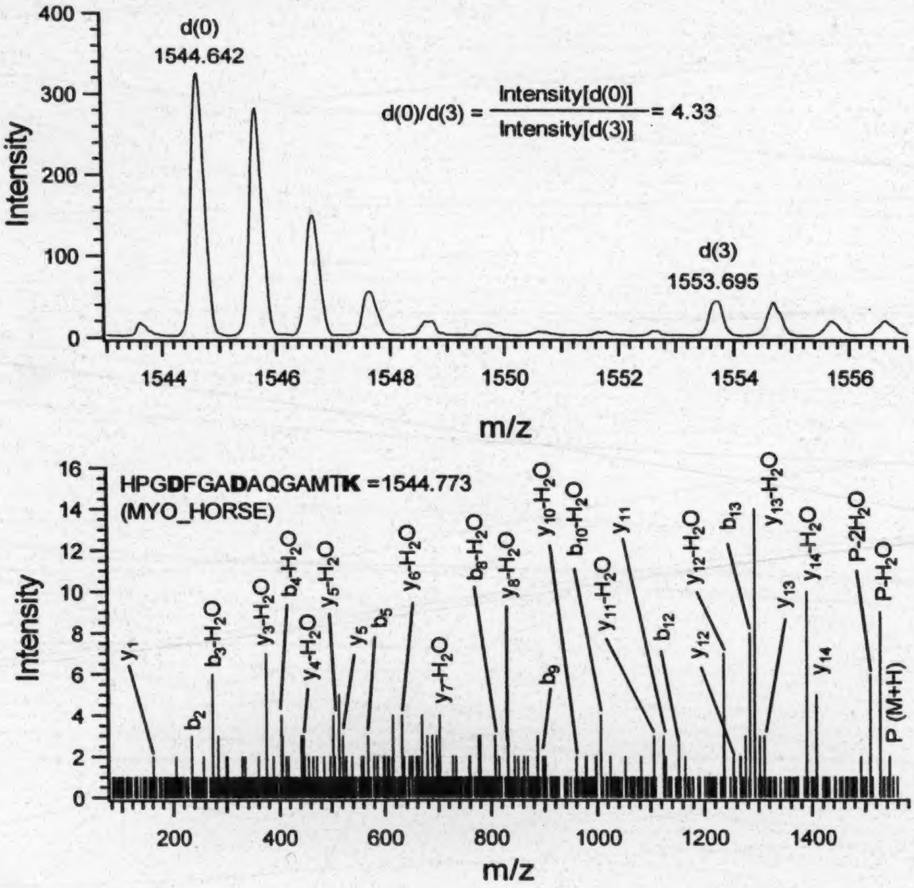

Figure 3.2 Quantification and identification of a peptide generated from a mixture of standard proteins. (A) Expanded MALDI mass spectrum. The relative quantities of the protein in the d(0)- and d(3)-labeled mixtures were obtained by comparing the peak intensities of the monoisotopic peaks at m/z values of 1544.642 and 1553.695, respectively. (B) MALDI MS/MS spectrum of the peptide ion at m/z 1544.642. The spectrum obtained was matched to the sequence HPGDFGADAQGAMTK (calculated m/z 1544.773) from MYG_HORSE, where the carboxylic acid groups in D and C-terminus were modified with the d(0)- form of the methanol reagent.

In addition, no fragmentation from the labeling group itself was observed in the MS/MS spectra of esterification and dimethylation labeled peptides as shown in the example of Figure 3.2B.

The results in Figure 3.2A show the manner in which peptide quantification can be obtained for the peptide pair without a significant overlap of the isotope envelopes. In this case, the relative ratios of the monoisotopic peaks of the d(0)- and d(3)-esterification labeled peptide pairs can be readily determined. The predominance of single-charged molecular ion peaks in MALDI analysis, along with the high resolution and high mass accuracy of MALDI QqTOF MS instrument, makes this measurement straightforward. Peak pairs separated by mass difference of $\Delta = 3.018n$ in the MS spectrum can be selected, where n is the number of methyl esterification site in the peptide (one for the C terminus and one for each aspartic and glutamic acid in the peptide) and 3.018 is the mass difference between the d(0)- and d(3)-methyl group. When dimethylation labeling is used, mass difference of $4.024m$ should be used to select peak pairs, where m is the dimethylation labeling site(s) in the peptide sequence (one for N terminus and one for lysine in the peptide).

As shown in Tables 3.1 and 3.2, for the peptides identified from this control mixture, the maximum error between the observed and expected d(0)/d(3) values is 4.0% and the maximum error between the observed and expected d(0)/d(2) values is 2.5%, indicating that the relative quantities of proteins are accurately measured using the LC-MALDI approach described herein.

In Tables 3.1 and 3.2, every protein listed has been identified and quantified based on several differential esterification or dimethylation labeled pairs, which increase the accuracy of the measurement. Comparing the data listed in Tables 3.1 and 3.2, it is found that both the accuracy and precision are improved using dimethylation labeling, compared to esterification labeling. This may be due to the different isotope effect between two labeling methods (see below). In addition, deamidation of glutamine and asparagine residues was also observed and these side-reactions could not be readily controlled, since they were not to completion during the labeling experiments, which could affect the relative abundance of the intended peptide pairs. However, no side reaction product was observed for dimethylation labeling.

For larger peptides where the isotope envelope of the molecular ions is broad, separation of 3 or 4 Da between the peptide pair may become a challenge for accurate determination of their intensity ratio. Partial overlap of isotopic envelopes from a peptide pair was observed for larger peptides (e.g., those with masses of greater than 2000 Da). However, as the peptide mass increases, the chance of multiple labeling of the peptide increases. Thus only a small portion of the peptide pairs at the high mass region contain a single labeling group. To eliminate the effect of peak overlap on quantification, the following strategy was used to determine the peak ratio between a peptide pair. An example is shown in Figure 3.3 where the overlapped peptide pair is de-convoluted by using the MS-isotope program, an online tool for calculating and visualizing isotope patterns of peptides from the UCSF Mass Spectrometry Facility

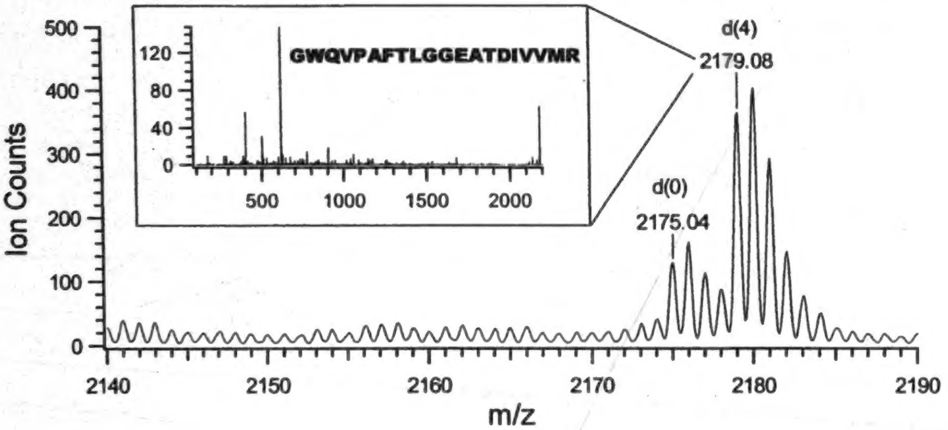

Figure 3.3 Quantification and identification of a peptide pair with significant overlap of isotope envelops. The d(0)/d(4) value was determined by comparing the corrected signal intensities of monoisotopic peaks of at m/z of 2175.04 and 2179.08 after the deconvolution and background subtraction. The d(4)-labeled peptide at m/z of 2179.08 was chosen for CID and successfully identified as GWQVPAFTLGGEATDIVVMR.

(http://prospector.ucsf.edu/ucsfhtml4.0/msiso.htm). Initially, the contribution of background to the monoisotopic peak intensities of the peptide pair was corrected using the observed monoisotopic peak intensities of the pair subtracted from the background. Next, the contribution of the overlapped isotope envelopes to the observed monoisotopic peak intensity of d(4)-GWQVPAFTLGGEATDIVVMR was calculated from the observed monoisotopic peak intensity of d(0)-GWQVPAFTLGGEATDIVVMR multiplied by a ratio factor determined from the expected relative intensities of isotope peaks of the peptide molecular ion. Third, the corrected monoisotopic peak intensity of d(4)-GWQVPAFTLGGEATDIVVMR was calculated using the observed monoisotopic peak intensity of d(4)-GWQVPAFTLGGEATDIVVMR subtracted by the calculated contribution from the overlapped isotope pattern and the background. Finally, the ratio between this peptide pair was calculated using the observed monoisotopic peak intensities of d(0)- and d(4)-GWQVPAFTLGGEATDIVVMR. While this method of calculation is a bit involved, it does provide a more accurate analysis of the peak pairs. The results shown in Tables 3.1 and 3.2 illustrate that accurate quantitative data can be obtained even for larger peptides.

3.3.2 Isotope Effect

The use of deuterium as the stable isotope label results in a primary isotope effect that can cause differential elution between the deuterated and nondeuterated peptides during μLC which may compromise measurement precision and accuracy [41-43]. In this study, the isotope effect resulting from differential labeling of peptides with d(0)- or d(3)-methanol and d(0)- or d(2)-formaldehyde has been examined. Myoglobin was digested to peptides with trypsin and the solution was divided into four equal aliquots, two of which were esterified separately using d(0)- or d(3)-methanol and another two of which were dimethylated separately using d(0)- or d(2)-formaldehyde. After labeling, the d(0)- and d(3)-methylated myoglobin digests or d(0)- and d(2)-formaldehyde labeled myoglobin digests were combined and the mixtures were separated by microbore RP-LC. The elute was analyzed by ESI MS. Reconstructed ion chromatograms of differentially labeled peptide pairs were obtained using Bruker Daltonics DataAnalysis software. The time shift between the differential esterification labeled pair or dimethylation labeled pair was calculated by comparison of elution time at the peaks of extracted ion chromatograms of the corresponding pair.

I have also observed that the d(3)-methyl esterification labeled peptide or d(2)-dimethylation labeled peptide elutes earlier than its counterpart, while the isotopic effect associated with dimethylation labeling (~1 or 2 s) is much smaller compared to that associated with the esterification labeling (~7 s). Thus, obtaining the correct quantitative peak ratios of the peptide pairs become difficult if the d(0)- and d(3)-methyl esterification labeled peptides are analyzed by LC-ESI MS. However, in microbore LC-MALDI, LC fractions were directly collected onto a MALDI sample plate in 1-min time intervals followed by MALDI QqTOF analysis. The isotope effect can be reduced. In this work, the retention time difference was accounted by scanning for the peptides which could be observed in MALDI MS spectra of consecutive fractions. Then the peptides' signal intensities were summed from adjacent spots to determine an accurate abundance ratio over the entire peptide elution profile. This way of data analysis is proved to be useful as demonstrated by the quantification results listed in Tables 3.1 and 3.2. The observed d(0)/d(3) or d(0)/d(2) ratio matches well with the expected one.

However, it was later found that, in a more complicated sample such as *E. coli* cell extract analysis, quantification could become difficult for some peptides with the esterification labeling method. The difference between the ratios of differential esterification labeled peptide pair in consecutive fractions could become very large. This difficulty is not encountered (see below) when using the differential dimethylation labeling method, because the isotopic effect associated with it is much smaller compared to that associated with the esterification labeling. The smaller isotope effect may be due to the factor that, in dimethylation labeling, the deuterium is associated with the charged amino residues which do not interact with the RP-LC stationary phase, thereby resulting in less chromatographic separation between the labeled and unlabeled peptides [43].

While good quantitative results can be obtained using esterification labeling for simple protein mixtures such as that shown in Table 3.1, after working with more complex samples, it was found that this labeling method had the following limitations. First, partial deamidation of glutamine and asparagine residues will increase the complexity of the peptide sample mixture. Second, anhydrous reaction conditions are difficult to obtain, resulting in incomplete labeling of the peptides. This is particularly true for the peptides with more than three labeling sites (data not shown) even after the labeling process is repeated. Finally, esters can be partially hydrolyzed by acidic mobile phase. The hydrolysis process is facilitated in the approach described herein since a heated interface is used. This labeling method is not compatible with separation methods using acidic mobile phases such as in isoelectric focusing chromatography. Therefore, the approach based on stable isotopic esterification labeling will be limited to the quantification of relatively simple protein mixtures. On the other hand, dimethylation labeling is performed under mild reaction condition with no apparent side reaction and the resultant peptides are stable, which makes this labeling method more compatible with microbore LC-MALDI using the heated droplet interface in analyzing complicated peptide mixtures. Small isotope effect associated with dimethylation labeling is another advantage, which helps in obtaining good precision and accuracy.

3.3.3 Quantification of Low Abundance Proteins

In many proteomics applications, protein concentrations in a mixture can vary in the order of 10^4 or higher. Quantitative analysis of low abundance proteins in a protein mixture is important and often requires an extensive pre-fractionation to reduce the concentration dynamic range so that the final mixture can be properly handled by MS. Proteins can be directly fractionated or can be converted into peptides followed by peptide fractionation. If an MS technique can only detect and quantitate a peptide mixture with a limited concentration dynamic range, an extensive pre-fractionation is required to reduce the complexity of the sample, and even so, there is a great risk of not detecting the peptides from the low abundance proteins, because of the final mixture after a limited number of separation steps can still be quite complex. One important feature of the microbore LC-MALDI based approach to protein quantification and identification is the ability to analyze a protein mixture with a relatively wider range of concentration differences, compared to μLC-based MS methods. This is due to the possibility of injecting a larger amount of samples to a microbore column, compared to a capillary column. The amount limit of injection is proportional to the square of column i.d..

To gauge the performance of the microbore LC-MALDI approach for analyzing a mixture of low and high abundance proteins, a set of experiments were carried out using a mixture of two standard proteins with varying amount ratios. As described in the Experimental section, the mixture consisted of a large amount of ubiquitin and a small amount of myoglobin. Figure 3.4A-D shows quantification and identification of d(0)-FDKFK and d(12)-FDKFK, a peptide pair from myoglobin. In this case, a total of 7 μg of peptide sample was injected on the column-this amount corresponds to the manufacturer's specified capacity for the column. If the digestion efficiency of 100% was assumed, the amount of each peptide would be 20 fmol. Another peptide pair, d(0)-ALELFR and d(4)-ALELFR, from myoglobin was quantified and identified. This example illustrates that, using the 1.0 × 150 mm (microbore) C_{18} column, two peptide pairs resulting from myoglobin could be quantified and identified in the presence of many peptide pairs resulting from 10^4 excess of ubiquitin.

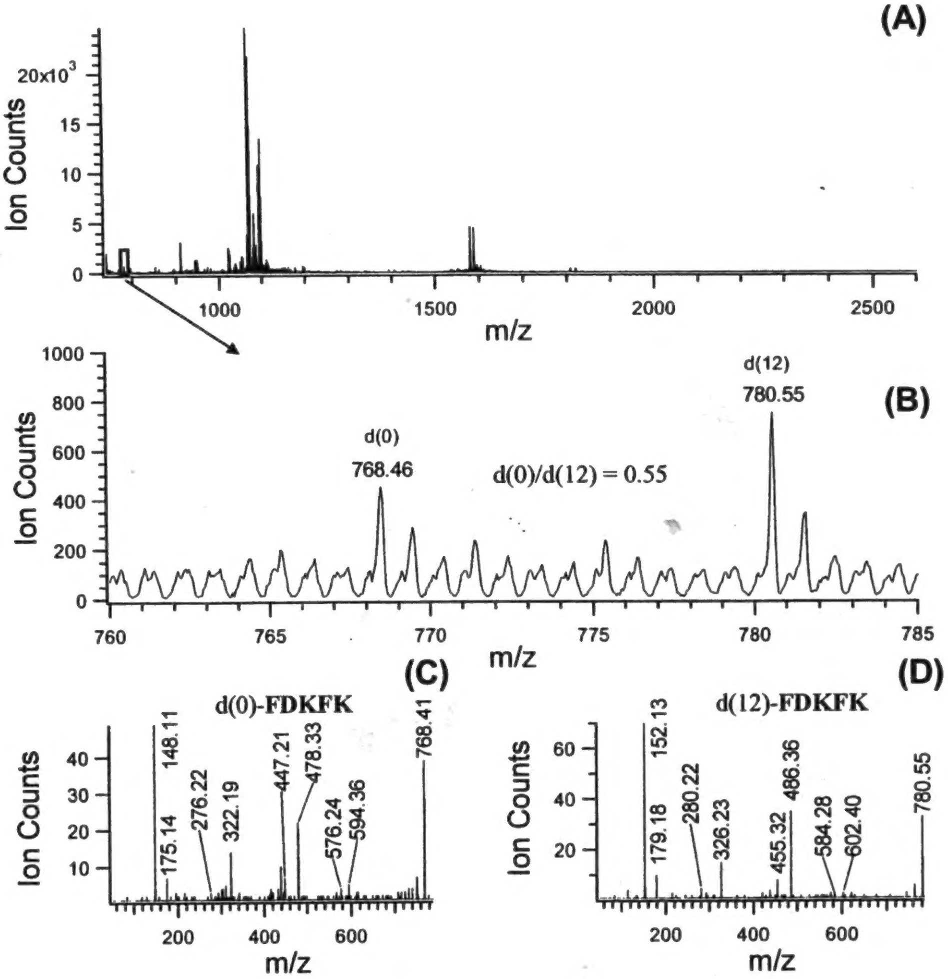

Figure 3.4 Quantification and identification of a low abundance protein. (A) The MALDI mass spectrum obtained for the fraction at 18 min from the microbore column separation. (B) The expended view indicated by rectangular in (A). (C) and (D) MALDI MS/MS spectra for a pair of peptides at m/z 768.461 and 780.553.

It should be noted that, if a capillary column (e.g., 0.3-mm i.d.) were used for analyzing the same sample, the sample had to be diluted by 10-fold so that it would not saturate the column. Of course, dilution of the sample would result in the reduction of the total amount injected to the column as well as the amount of the low abundance protein injected. In this case, the amount of the low abundance protein would be 2 fmol (assuming dilution of 10-fold), which is below the quantitative detection limit of current MALDI techniques.

3.3.4 *E. coli* Whole Cell Extracts

To demonstrate the applicability of the combination of differential dimethylation labeling and microbore LC-MALDI QqTOF MS to quantitatively analyze complex protein mixtures, this approach was applied to the analysis of *E. coli* whole cell extract digests. Protein extraction and digestion from *E. coli* were described in the Experimental Section. The resulting tryptic digest was divided into two vials with the amount ratio of 1:3. The peptides in the two vials were dimethylated with d(0)- or d(2)-formaldehyde, respectively. The labeled peptides were then combined and the mixture (about 160 μg) was loaded on the 1×150 mm C_{18} column and collected from 1 to 100 min at 1 min/well onto a MALDI plate.

Almost all peptides were detected from spots fractionated from 20 to 100 min. Among them, most peptides were detected only once from fractionated spots. However,

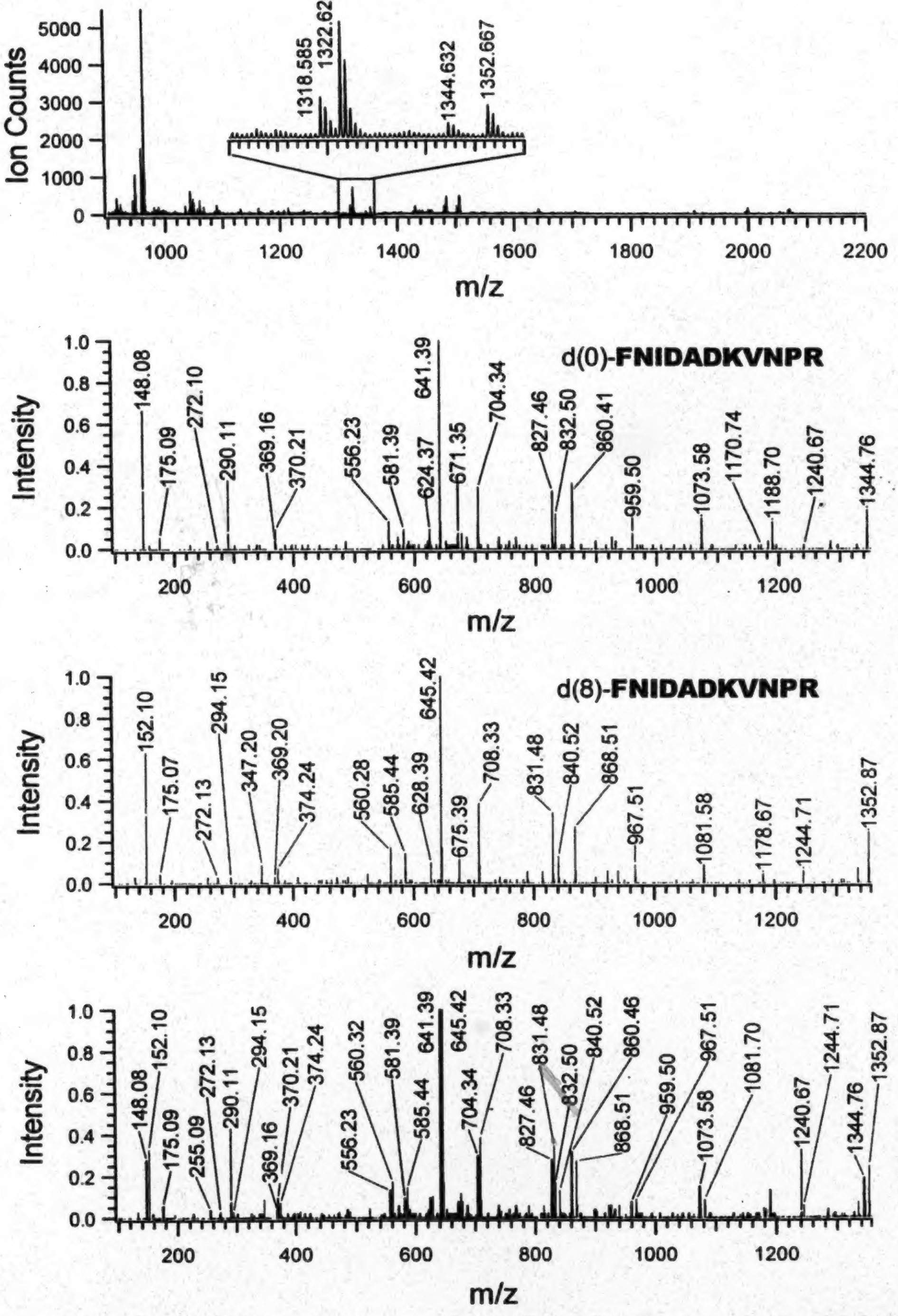

Figure 3.5 (A) MALDI mass spectrum obtained for the fraction at 45 min from the microbore column separation. (B) and (C) MALDI MS/MS spectra for a pair of peptides at m/z 1344.597 and 1352.667. (D) Overlaid MS/MS spectra of (B) and (C).

some peptides were observed in consecutive multiple fractions. To get accurate ratios for those peptides present in consecutive fractions, the peak intensities of monoisotopic peaks in each fraction were summed and the summed background was subtracted. The final ratio was calculated using summed and corrected intensities for each pair.

Figure 3.5A shows an example of LC fraction at 45 min being characterized by the MALDI QqTOF mass spectrometer. Two representative MS/MS spectra of a d(0)- and d(2)-formaldehyde-labeled peptide pair are shown in Figures 3.5B and C. In Figure 3.5A, 12 peptide pairs are quantified and selected for MS/MS analysis, and database searchable MS/MS spectra are obtained from 11 peptide pairs. From these MS/MS spectra, 10 unique proteins are identified and the identification results are confirmed by comparison of MS/MS spectra of the d(0)- and d(2)-formaldehyde-labeled peptide pairs. An example of using the labeled peptide pairs to confirm peptide sequence results generated from database search is shown in Figure 3.5D. This is one of the advantages of using dimethylation labeling where the labeled amine groups provide a mass tag that assists in MS/MS spectral interpretation.

Table 3.3 summarizes the results of quantification and identification from fraction 45. The observed 3.0% error indicates that the relative protein quantities were accurately measured. The relative standard deviation is about 8.8%, which is well within 15% that is widely used as a reproducibility or precision threshold for bioassay.

From this one-dimensional LC-MALDI experiment, 423 peptide pairs, which lead to the identification of 160 proteins, were quantified with d(0)/d(2) values having a range

Table 3.3 Proteins identified and quantified in HPLC fraction 45 of *E. coli* extract samples.

m/z	Peptide sequence identified	Protein	Intensity (counts)	Abundance ratio*			
				Obs'd	Expt'd	%SD	%Err
943.46	WTGIPVSR	CLPB_ECOLI	347	0.34	0.33	8.8	3.0
947.48			1029				
958.39	GYRPQFY	EFTU_ECOLI	1727	0.32			
962.42			5464				
1034.43	NIYDYYK	TALA_ECOLI	197	0.33			
1042.48			591				
1055.51	NQYYGITAGPAYR	MASY_ECOLI	136	0.39			
1059.55			342				
1085.52	TAIVEGLAQR	CLPB_ECOLI	68	0.32			
1089.54			215				
1318.59	AYGSTNPINVVR	RS5_ECOLI	227	0.31			
1322.62			698				
1344.60	FNIDADKVNPR	ODP1_ECOLI	68	0.39			
1352.67			177				
1479.64	ITPTFTEESDGVR	YFCZ_ECOLI	162	0.39			
1483.64			414				
1501.62	NQYYGITAGPAYR	OMPX_ECOLI	189	0.41			
1505.65			459				
1896.00	HQKPVPALNQPGGIVEK	RL24_ECO57	28	0.34			
1908.10			83				
1993.81	not identified		49	0.33			
1997.81	not identified		147				
2057.84	AQGNMPAYGYTPPYTDGAK	OPPA_ECOLI	30	0.31			
2065.89			98				

*Obs'd: observed abundance ratio from the peptide pair; Exp'd: expected abundance ratio; %SD: relative standard deviation of the observed abundance ratios; %Err: percent error calculated from the average of the observed abundance ratios and the expected ratio.

from 0.26 to 0.42, an average of 0.34 and a relative standard deviation of 10.3% (see Table 3.4). The quantitative errors at the protein level are less than 24%. For proteins quantified by using two or more peptides (80 out of 160 proteins), the error is less than 13%. This example illustrates that the microbore LC-MALDI method combined with dimethylation labeling can be used to provide reliable quantitative information on relative abundance of complex protein samples. It also demonstrates that, while positive protein identification can be obtained based on a single peptide MS/MS data using overlaid spectral information generated from the differential peptide pair, better quantitative accuracy is obtained by averaging the abundance ratios of multiple peptide pairs from a protein. With a larger sample loading afforded by microbore LC-MALDI, multiple peptides from a protein can be more readily detected, compared to capillary LC-MS, thereby increasing the overall quantitative accuracy of the experiment. In the future, the addition of other extensive sample pre-fractionation before microbore LC-MALDI should reduce the complexity of the peptide mixture and result in much more comprehensive detection and quantification of biologically interesting proteins from complicated samples.

3.4 Conclusions

An approach of quantitative proteome analysis based on microbore LC-MALDI MS have been presented. It has been shown that dimethylation labeling is more compatible with microbore LC-MALDI using a heated droplet interface than esterification labeling. Some of the important features of the developed approach can be summarized as followings. First of all, dimethylation labeling is inexpensive and can be done with commercially available reagents. And it has high reaction efficiency, no side reaction, and small isotope effect on reversed-phase separation. Second, the overlaid fragment ion spectra generated from a pair of differentially labeled peptides can be used to confirm peptide sequences obtained from MS/MS database search. Third, microbore LC-MALDI can facilitate the detection and quantification of low-abundance proteins in complex protein mixtures since the high loading capacity is allowed by using the heated droplet interface. Fourth, abundance ratio of peptides in two samples can be used as a guide for target protein identification as in all LC-MALDI methods. Finally, dimethylation labeling is a global proteome analysis method. Thus, the combination of dimethylation labeling with microbore LC-MALDI should be applicable for comprehensive quantitative proteome analysis.

3.5 Literature Cited

(1) Tao, W. A.; Aebersold, R. *Curr. Opin. Biotechnol.* **2003**, *14*, 110-118.

(2) Duggan, D. J.; Bittner, M.; Chen, Y.; Meltzer, P.; Trent, J. M. *Nat. Genet.* **1999**, *21*, 10-14.

(3) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

(4) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. E. *J.*

Chromatogr. B Biomed. Sci. Appl. **2000**, *745*, 197-210.

(5) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836-2842.

(6) Qiu, Y.; Sousa, E. A.; Hewick, R. M.; Wang, J. H. *Anal. Chem.* **2002**, *74*, 4969-4979.

(7) Ong, S.-E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell. Proteomics* **2002**, *1*, 376-386.

(8) Griffin, T. J.; Han, D. K.; Gygi, S. P.; Rist, B.; Lee, H.; Aebersold, R.; Parker, K. C. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1238-1246.

(9) Griffin, T. J.; Gygi, S. P.; Ideker, T.; Rist, B.; Eng, J.; Hood, L.; Aebersold, R. *Mol. Cell. Proteomics* **2002**, *1*, 323-333.

(10) Gygi, S. P.; Rist, B.; Griffin, T. J.; Eng, J.; Aebersold, R. *J. Proteome Res.* **2002**, *1*, 47-54.

(11) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946-951.

(12) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, *20*, 512-515.

(13) Wang, S.; Regnier, F. E. *J. Chromatogr. A* **2001**, *924*, 345-357.

(14) Spahr, C. S.; Susin, S. A.; Bures, E. J.; Robinson, J. H.; Davis, M. T.; McGinley, M. D.; Kroemer, G; Patterson, S. D. *Electrophoresis* **2000**, *21*, 1635-1650.

(15) Geng, M.; Ji, J.; Regnier, F. E. *J. Chromatogr. A* **2000**, *870*, 295-313.

(16) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J.

K.; Haller, P.-V.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1214-1221.

(17) Hsu, J.; Huang, S.; Chow, N.; Chen, S. *Anal. Chem.* **2003**, *75*, 6843-6852.

(18) Everley, P. A.; Krijgsveld, J.; Zetter, B. R.; Gygi, S. P. *Mol. Cell. Proteomics* **2004**, *3*, 729-735.

(19) Che, F. Y.; Fricker, L. D. *Anal. Chem.* **2002**, *74*, 3190-3198.

(20) Krijgsveld, J.; Ketting, R. F.; Mahmoudi, T.; Johansen, J.; Artal-Sanz, M.; Verrijzer, C. P.; Plasterk, R. H. A.; Heck, A. J. R. *Nat. Biotechnol.* **2003**, *21*, 927-931.

(21) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., III. *Anal. Chem.* **2002**, *74*, 1650-1657.

(22) Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. *Anal. Chem.* **2003**, *75*, 1316-1324.

(23) Jiang, H.; English, A. M. *J. Proteome Res.* **2002**, *1*, 345-350.

(24) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163-170.

(25) Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem.* **2004**, *76*, 2748-2755.

(26) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6591-6596.

(27) Staes, A.; Demol, H.; Van Damme, J.; Martens, L.; Vandekerckhove, J.; Gevaert, K. *J. Proteome Res.* **2004**, *3*, 786-791.

(28) Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.;

Anderson, D. J.; Lipton, M. S.; Pasa-Tolic, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 2132-2149.

(29) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Tolic, L. P.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513-523.

(30) Beardsley, R. L.; Reilly, J. P. *J. Proteome Res.* **2003**, *2*, 15-21.

(31) Griffin, T. J.; Gygi, S. P.; Rist, B.; Aebersold, R.; Loboda, A.; Jilkine, A.; Ens, W.; Standing, K. G. *Anal. Chem.* **2001**, *73*, 978-986.

(32) Griffin, T. J.; Lock, C. M.; Li, X.-j.; Patel, A.; Chervetsova, I.; Lee, H.; Wright, M. E.; Ranish, J. A.; Chen, S. S.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 867-874.

(33) Zappacosta, F.; Annan, R. S. *Anal. Chem.* **2004**, *76*, 6618-6627.

(34) Zhang, B.; McDonald, C. M.; Li, L. *Anal. Chem.* **2004**, *76*, 992-1001.

(35) Zhang, N.; Li, N.; Li, L. *J. Proteome Res.* **2004**, *3*, 719-727.

(36) Li, N.; Shaw, A. R. E.; Zhang, N.; Mak, A.; Li, L. *Proteomics* **2004**, *4*, 3156-3166.

(37) Hunt, D. F.; Yates, J. R., III. ; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 6233-6237.

(38) Means, G. E.; Feeney, R. E. *Biochemistry* **1968**, *7*, 2192-2201.

(39) Friedman, M.; Williams, L. D.; Masri, M. S. *Int. J. Pept. Protein Res.* **1974**, *6*, 183-185.

(40) Quach, T. T. T.; Li, N.; Richards, D. P.; Zheng, J.; Keller, B. O.; Li, L. *J. Proteome Res.* **2003**, *2*, 543-552.

(41) Zhang, R.; Regnier, F. E. *J. Proteome Res.* **2002**, *1*, 139-147.

(42) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. *Anal. Chem.* **2002**, *74*, 3662-3669.

(43) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. *Anal. Chem.* **2001**, *73*, 5142-5149.

Table 3.4 Proteins identified and quantified in *E.coli* extract samples.

#	Protein Name	Access ID	Unique Peptide Sequence*	Abundance Ratio**				
				Mea	Ave	Exp'd	%SD	%Err
1	Elongation factor Tu	P02990	TYGGAAR	0.33	0.34	0.33	10.6	2.9
			LLDEGR	0.39				
			HTPFFK	0.35				
			GYRPQFY	0.32				
			REEIER	0.34				
			EHILLGR	0.37				
			LLPIEDVF	0.35				
			AGENVGVLLR	0.32				
			VGEEVEIVGIK	0.39				
			DFPGDDTPIVR	0.35				
			IHPIAMDDGLR	0.39				
			TTLTAAITTVLAK	0.31				
			AIDKPFLLPIEDVF	0.28				
			GQVLAKPGTIKPHTK	0.39				
			GITINTSHVEYDTPTR	0.32				
			TKPHVNVGTIGHVDHGK	0.35				
			ILELAGFLDSYIPEPER	0.32				
			ELLSQYDFPGDDTPIVR	0.26				
			AIDKPFLLPIEDVFSISGR	0.37				
2	Formate acetyltransferase 1	P09373	LHTYR	0.33	0.34	0.33	9	1.8
			IFTEYR	0.37				
			ALIPFGGIK	0.35				
			QMQFFGAR	0.34				

			LATAWEGFTK	0.39				
			GDWQNEVNVR	0.32				
			VDDLAVDLVER	0.33				
			LREEIAEQHR	0.32				
			SGVLTGLPDAYGR	0.3				
			TSTFLDVYIER	0.33				
			IVGLQTEAPLKR	0.36				
			AGAPFGPGANPMHGR	0.3				
			KSGVLTGLPDAYGR	0.32				
			RAGAPFGPGANPMHGR	0.38				
			ITEQEAQEMVDHLVMK	0.3				
			EMLLDAMENPEKYPQLTIR	0.39				
3	Glutamate decarboxylase alpha	P80063	LLTDFR	0.39	0.33	0.33	0	6.6
			YLSDHPK	0.27				
			YWDVELR	0.33				
			RFPLHEMR	0.36				
			EIPMRPGQLF	0.33				
			LQGIAQQNSFK	0.38				
			QIINDELYLDGNAR	0.29				
			EIPMRPGQLFMDPK	0.37				
			GFEMDFAELLLEDYK	0.32				
			NWIDKEEYPQSAAIDLR	0.35				
			GWQVPAFTLGGEATDIVVMR	0.31				
			DDVAFQIINDELYLDGNAR	0.28				
4	Isocitrate lyase	P05313	ADQIQW	0.3	0.33	0.33	10.5	0.5
			INNTFR	0.34				

			EWTQPR	0.33				
			TSEGFFR	0.42				
			SAGIEPGDPR	0.35				
			RFAQAIHAK	0.32				
			VQQPEFAAAK	0.36				
			THAGIEQAISR	0.33				
			VLVPTQEAIQK	0.35				
			RADQIQWSAGIEPGDPR	0.33				
			FLPIVADAEAGFGGVLNAF	0.28				
			DGYTFVSHQQEVGTGYFDK	0.31				
5	Outer membrane protein A	P02934	SLGVSYR	0.42	0.34	0.33	11	1.6
			LGGMVWR	0.31				
			AITPEIATR	0.36				
			DNTWYTGAK	0.34				
			EMGYDWLGR	0.37				
			DGSVVVLGYTDR	0.32				
			IGSDAYNQGLSER	0.35				
			VGFEMGYDWLGR	0.35				
			NHDTGVSPVFAGGVEY	0.29				
			ATLKPEGQAALDQLY	0.3				
			FGQGEAAPVVAPAPAPAPEVQTK	0.32				
6	DNA protection during starvation protein	P27430	ATVELLNR	0.34	0.33	0.33	10.7	0.5
			QAHWNMR	0.37				
			DLDKFLW	0.32				
			YAIVANDVR	0.35				
			IAVHEMLDGFR	0.4				

			TALIDHLDTMAER	0.34				
			QVIQFIDLSLITK+Pyro-glu(N-term Q)	0.29				
			QVIQFIDLSLITK	0.3				
			GANFIAVHEMLDGFR	0.35				
			SYPLDIHNVQDHLK	0.31				
			AVQLGGVALGTTQVINSK	0.28				
7	Malate synthase A	P08997	FSQGR	0.36	0.34	0.33	7.9	2.3
			VIDGQINLR	0.39				
			KNQLEVMR	0.33				
			VIASELGEER	0.33				
			TRPYGEQEK	0.35				
			RVEITGPVER	0.36				
			AFTRPYGEQEK	0.33				
			LTELVTHFTPQR	0.36				
			VFMADFEDSLAPDWNK	0.29				
			TEQATTTDELAFTRPYGEQEK	0.34				
			LMEQITTSDELIDFLTLPGYR	0.31				
8	Lysyl-tRNA synthetase, heat inducible	P14825	SWGLGR	0.29	0.34	0.33	10.7	2.6
			VFEINR	0.4				
			LVVGGFER	0.35				
			WDLGDIIGAR	0.39				
			ASFVTLQDVGGR	0.32				
			DVILFPAMRPQK	0.34				
			GANEAIDFNDELR	0.38				
			ADYHDLIELTESLFR	0.33				
			ALRPLPDKFHGLQDQEVR	0.34				

			MAYADYHDLIELTESLFR	0.3				
9	Tagatose-1,6-bisphosphate aldolase gatY	P37192	AQNISR	0.33	0.33	0.33	7.6	1.5
			TNPAQAR	0.33				
			DYLQSAK	0.37				
			NAFSQALK	0.29				
			ASAPALDFSR	0.3				
			NYLTEHPEATDPR	0.34				
			THAGTENLLALVSAMAK	0.32				
			QYHHPLAIHLDHHTK+Pyro-glu(N-term Q)	0.34				
			QYHHPLAIHLDHHTK	0.35				
			EFAEATGIDSLAVAIGTAHGMY	0.31				
10	Protein yfiD	P33633	LGDIEYR	0.37	0.35	0.33	9.6	5.8
			LLDSEKGEAR	0.36				
			FNSLTPEQQR	0.38				
			AANDDLLNSFW	0.36				
			EVPVEVKPEVR	0.34				
			AGYAEDEVVAVSK	0.27				
			HPEKYPQLTIR	0.34				
			VEGGQHLNVNVLR	0.39				
			ETLEDAVKHPEKYPQLTIR	0.36				
11	ClpB protein	P03815	TIQVLQR	0.35	0.33	0.33	5.3	0.3
			WTGIPVSR	0.34				
			TAIVEGLAQR	0.32				
			MEQELHHR	0.35				
			YRGEFEER	0.36				
			NILLQVLDDGR	0.32				

			GGESVNDQGAEDQR	0.32				
			VFVAEPSVEDTIAILR	0.33				
			LVGAPPGYVGYEEGGYLTEAVR	0.3				
12	Pyruvate dehydrogenase E1 component	P06958	AFLEGR	0.35	0.33	0.33	11.5	0.9
			LFAEQVR	0.32				
			LVPIIADEAR	0.32				
			LTQEQLDNFR	0.27				
			FNIDADKVNPR	0.39				
			DWLQAIESVIR	0.3				
			AQYLIDQLLAEAR	0.33				
			LELPSLQDFGALLEEQSK	0.36				
13	30S ribosomal protein S4	P02354	MGFGATR	0.33	0.31	0.33	7.4	5.5
			LDNVVYR	0.33				
			AALELAEQR	0.3				
			IEQAPGQHGAR	0.34				
			EKPTWLEVDAGK	0.28				
			GNTGENLLALLEGR	0.33				
			LKGNTGENLLALLEGR	0.29				
14	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	P31217	YGDEQVK	0.38	0.33	0.33	8.2	1.4
			WNETILPR	0.32				
			DDERYPGHDPR	0.35				
			HGESQWNKENR	0.34				
			FTGWYDVDLSEK	0.3				
			VIPYWNETILPR	0.31				
			NVLDELDQAWLPVEK	0.3				
15	Phosphoglycerate kinase	P11665	EFPNFR	0.39	0.34	0.33	7.6	3.2

			VMDAFGTAHR	0.33				
			ALKEPARPMVAIVGGSK	0.33				
			SIAGGGDTLAAIDLFGIADK	0.32				
			VMVTSHLGRPTEGEYNEEF	0.36				
			ADEQILDIGDASAQELAEILK	0.33				
16	Glyceraldehyde 3-phosphate dehydrogenase A	P06977	YDSTHGR	0.35	0.36	0.33	10.8	8.6
			VTAERDPANLK	0.4				
			TVDGPSHKDWR	0.34				
			VPTPNVSVVDLTVR	0.39				
			RVPTPNVSVVDLTVR	0.3				
			LVSWYDNETGYSNK	0.39				
17	60 kDa chaperonin	P06139	APGFGDR	0.33	0.35	0.33	8.1	6.5
			AMEAPLR	0.35				
			QQIEEATSDYDREK+Pyro-glu(N-term Q)	0.4				
			QQIEEATSDYDR	0.38				
			FINKPETGAVELESPF	0.33				
			DTTTIIDGVGEEAAIQGR	0.35				
18	50S ribosomal protein L6	P02390	RPEPYK	0.37	0.34	0.33	13.1	2.0
			YADEVVR	0.34				
			KLQLVGVGYR	0.32				
			HADNTLTFGPR	0.27				
			DGYADGWAQAGTAR	0.35				
			ALLNSMVIGVTEGFTK	0.4				
19	Lysyl-tRNA synthetase	P13030	SWGLGR	0.29	0.32	0.33	14.3	3.2
			VFEINR	0.39				
			LVVGGFER	0.35				

			ASFVTLQDVGGR	0.32				
			DLIELTESLFR	0.27				
20	30S ribosomal protein S1	P02349	AISLSVR	0.35	0.33	0.33	9.4	2.0
			AFLPGSLVDVRPVR	0.33				
			TESFAQLFEESLK	0.29				
			MTESFAQLFEESLK	0.36				
			VVNVGDVVEVMVLDIDEER	0.3				
21	Tryptophanase	Q8XB34	HLPEPFR	0.34	0.35	0.33	3.3	6.5
			GDEAYSGSR	0.35				
			AYREEAIIK	0.36				
			GNFDLEGLER	0.37				
			DWTIEQITR	0.36				
22	50S ribosomal protein L5	P02389	DFPFR	0.39	0.35	0.33	11.5	3.7
			LITIAVPR	0.32				
			NYNSVMQVPR	0.39				
			ALLAAFDFPFRK	0.32				
			EQIIFPEIDYDKVDR	0.32				
23	50S ribosomal protein L3	P60438	GLWEFR	0.32	0.31	0.33	5.7	7.4
			DLANDGYR	0.33				
			VTVQSLDVVR	0.3				
			IFTEDGVSIPVTVIEVEANR	0.29				
24	Alkyl hydroperoxide reductase subunit C	P26427	EDEGLADR	0.3	0.32	0.33	11.2	3.1
			YAMIGDPTGALTR	0.38				
			VVDPQGIIQAIEVTAEGIGR	0.3				
			ATFVVDPQGIIQAIEVTAEGIGR	0.31				
25	Bacterioferritin	P11056	LNIGEDVEEMLR	0.39	0.36	0.33	9.4	6.7
			MGLQNYLQAQIR	0.31				

			EAIGYADSVHDYVSR	0.37				
			ILFLEGLPNLQDLGK	0.36				
26	Outer membrane protein slp	P37194	GNNQPDIQK	0.33	0.33	0.33	9.7	0.2
			QSGFLDPVNYR	0.38				
			SFVAVHNQPGLY	0.32				
			AKPDIEANYQGR	0.3				
27	30S ribosomal protein S9	P02363	SLEQYFGR	0.4	0.36	0.33	8.9	6.6
			GGGISGQAGAIR	0.34				
			AENQYYGTGR	0.32				
			ALMEYDESLR	0.36				
28	50S ribosomal protein L1	P02384	KSDQNVR	0.4	0.34	0.33	11.5	2.3
			VVGQLGQVLGPR	0.35				
			ENLEALLVALK	0.31				
			QYDINEAIALLK+Pyro-glu(N-term Q)	0.32				
29	30S ribosomal protein S16	P02372	MVTIR+Oxidation(M)	0.35	0.33	0.33	9.2	0.4
			VGFFNPIASEK	0.36				
			IAHWVGQGATISDR	0.29				
			VGFFNPIASEKEEGTR	0.32				
30	Putative tagatose 6-phosphate kinase gatZ	P37191	YYWPHSR	0.37	0.33	0.33	8.2	0.6
			AQWIENTR	0.32				
			TGFNDSLLDIR	0.33				
			MVYEAHSTDYQTR	0.31				
31	Flavoprotein wrbA	P30849	FGNMSGQMR	0.35	0.3	0.33	11.2	8.6
			AAQELFDVSQVR	0.29				
			GGTPYGATTIAGGDGSR	0.27				
			RVPETMPPQLFEK	0.31				
32	50S ribosomal subunit protein L15	P02413	FGFTSR	0.29	0.33	0.33	14.2	2.1

			GFEGGQMPLYR	0.37				
			RGFEGGQMPLYR	0.29				
			VILAGEVTTPVTVR	0.36				
33	Elongation factor G	P02996	AGPLAGYPVVDMGIR	0.34	0.35	0.33	5.3	5.8
			IATDPFVGNLTFFR	0.34				
			EFNVEANVGKPQVAYR	0.37				
34	Outer membrane protein W	P21364	LGGAQQHDSVR	0.33	0.36	0.33	8.6	6.7
			MATDNIGVELLAATPFR	0.35				
			TYMATDNIGVELLAATPFR	0.39				
35	Cysteine synthase A	P11096	IAGVGTGGTLTGVSR	0.32	0.34	0.33	5.3	2.8
			IGANMIWDAEKR	0.34				
			IFEDNSLTIGHTPLVR	0.36				
36	Glutaredoxin 2	P39811	SPAIEEWLR	0.31	0.32	0.33	5.1	2.6
			SAFDEFSTPAAR	0.34				
			EASAGNFADLLAHSDGLIK	0.33				
37	Aldehyde dehydrogenase A	P25553	AQPEWEALPAIER	0.37	0.34	0.33	12	2.0
			GYYYPPTLLLDVR	0.39				
			GDAWIDVVNPATEAVISR	0.31				
38	30S ribosomal protein S5	P02356	VGFGYGK	0.33	0.35	0.33	14.2	4.6
			AYGSTNPINVVR	0.31				
			VFMQPASEGTGIIAGGAMR	0.41				
39	30S ribosomal protein S7	P02359	AFAHYR	0.35	0.37	0.33	5.5	11.5
			SALETLAQR	0.37				
			SELEAFEVALENVRPTVEVK	0.39				
40	50S ribosomal protein L19	P02420	LQAFEGVVIAIR	0.38	0.35	0.33	11.7	4.0
			QDVPSFRPGDTVEVK	0.36				
			QDVPSFRPGDTVEVK+Pyro-glu(N-term Q)	0.3				

41	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase	P07016	GLVTPVLR	0.37	0.36	0.33	6.9	9.0
			KQYGEAFEKR	0.39				
			ESAPAAAAPAAQPALAAR	0.34				
42	50S ribosomal protein L28	P02428	FVTLR	0.35	0.36	0.33	7	9.3
			FWVESEKR	0.35				
			FLPNLHSHR	0.39				
43	30S ribosomal protein S3	P02352	LGIVKPW	0.29	0.31	0.33	7.8	8.2
			VTIHTARPGIVIGK	0.3				
			KVVADIAGVPAQINIAEVR	0.33				
44	Fructose-bisphosphate aldolase class II	P11604	ANEAYLQGQLGNPK	0.39	0.37	0.33		10.1
			IFDFVKPGVITGDDVQK	0.35				
45	Enolase	P08324	DAGYTAVISHR	0.37	0.35	0.33		5.4
			AFTSEEFTHFLEELTK	0.33				
46	DNA-binding protein H-NS	P08936	REEESAAAAEVEER	0.29	0.31	0.33		7.4
			EMLIADGIDPNELLNSLAAVK	0.33				
47	Hypothetical protein yqjD	P42617	ADEYVR	0.38	0.33	0.33		0.2
			SKEHTTEHLR	0.29				
48	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	P03948	FADYDEAR	0.37	0.37	0.33		10.7
			MQQLQNIIETAFER	0.36				
49	DNA-binding protein HU-beta	P02341	ALDAIIASVTESLK	0.36	0.35	0.33		5.1
			TGRNPQTGK	0.34				
50	GrpE protein	P09372	VANLEAQLAEAQTR	0.42	0.37	0.33		11.7
			FINELLPVIDSLDR	0.33				
51	Chaperone protein dnaK	P04475	FQDEEVQR	0.37	0.37	0.33		11.7
			AKLESLVEDLVNR	0.38				

52	Protein yjbJ	P32691	EVVDWETR	0.33	0.35	0.33	3.6
			MNKDEAGGNWK	0.36			
53	50S ribosomal protein L4	P02388	SILSELVR	0.34	0.36	0.33	7.5
			AARPQDHSQK	0.37			
54	PTS system, glucose-specific IIA component	P08837	IAEEGQR	0.31	0.32	0.33	5.0
			LSGSVTVGETPVIR	0.32			
55	Protein yjgF	P39330	IEIEAIAVR	0.38	0.36	0.33	8.3
			TGEVPADVAAQAR	0.35			
56	50S ribosomal protein L10	P02408	LATLPTYEEAIAR	0.38	0.37	0.33	10.5
			AAAFEGELIPASQIDR	0.36			
57	Ferritin 1	P23887	YVSEQHEEEK	0.3	0.32	0.33	3.5
			LFDYLTDTGNLPR	0.34			
58	Unknown protein from 2D-page	P39169	IYPGQVLR	0.33	0.33	0.33	1.8
			LWDAVTGQHDKDDQAK	0.32			
59	DNA-binding protein HU-alpha	P02342	TGRNPQTGK	0.34	0.36	0.33	9.0
			AALESTLAAITESLK	0.36			
60	Purine nucleoside phosphorylase	P09743	THEQTTAAER	0.32	0.3	0.33	11.4
			YIAETFLEDAR	0.27			
61	Aldehyde-alcohol dehydrogenase	P17547	LLAWLETLK	0.35	0.35	0.33	5.3
			SIYFR	0.35			
62	Succinyl-CoA synthetase beta chain	P07460	DQSQEDPR	0.4	0.35	0.33	4.7
			AVLVNIFGGIVR	0.3			
63	ATP synthase beta chain	P00824	GQMNEPPGNR	0.37	0.35	0.33	5.3
			DVLLFVDNIYR	0.33			
64	Galactitol-1-phosphate 5-dehydrogenase	P37190	HQDEVR	0.28	0.29	0.33	12.2
			SSPWPGQEWETASR	0.31			

65	Malate dehydrogenase	P06994	FGLSLVR	0.35	0.31	0.33	8.0
			LFGVTTLDIIR	0.27			
66	50S ribosomal protein L7/L12	P02392	TEFDVILK	0.42	0.37	0.33	11.4
			FGVSAAAAVAVAAGPVEAAEEK	0.32			
67	30S ribosomal protein S21	P02379	AGVLAEVR	0.31	0.32	0.33	4.4
			EFYEKPTTER	0.32			
68	Hypothetical protein yegP	P76402	SNSPQEER	0.36	0.36	0.33	9.3
			AGWFELSK	0.37			
69	50S ribosomal protein L32	P02435	AVQQNKPTR	0.34	0.32	0.33	4.1
			HHITADGYYR	0.3			
70	Outer membrane protein X	P36546	NQYYGITAGPAYR	0.39	0.37	0.33	9.8
			YRYEEDNSPLGVIGSF	0.34			
71	Ribosome modulation factor	P22986	GYQAGIAGR	0.38	0.36	0.33	6.8
			SQWLGGWR	0.33			
72	Hydrogenase-1 small chain	P19928	HNQQPTETEHQPGNEDK	0.33	0.36	0.33	9.3
			RHNQQPTETEHQPGNEDK	0.4			
73	Superoxide dismutase [Fe]	P09157	SLEEIIR	0.41	0.36	0.33	8.4
			NARPGYLEHFW	0.32			
74	Transaldolase A	P78258	NIYDYYK	0.33	0.31	0.33	6.0
			LIPPSQTFPRPAPMSEAEFR	0.29			
75	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	P06959	QEAAPAAAPAPAAGVK+Pyro-glu (N-term Q)	0.29	0.32	0.33	3.3
			VPDIGADEVEITEILVK	0.34			
76	30S ribosomal protein S17	P02373	SWTLVR	0.26	0.30	0.33	10.2
			SIVVAIER	0.34			
77	50S ribosomal protein L9	P02418	LFGSIGTR	0.36	0.37	0.33	12.0

78	Universal stress protein G	P39177	NIEFFEAR	0.39	0.36	0.33	9.5
			RFEEHLQHEAQER	0.35			
			HANLPVLVVR	0.38			
79	50S ribosomal protein L20	P02421	GYYGAR	0.32	0.35	0.33	4.7
			QLWIAR	0.38			
80	Chaperone protein hchA (Hsp31)	P31658	VMPFFEQHK	0.32	0.32	0.33	4.8
			TSPVSDLDGVDYPKPYR	0.31			
81	ElaB protein	P52084	IDDDLTLLSETLEEVLR	0.34		0.33	0.9
82	DNA-directed RNA polymerase alpha chain	P00574	AATILAEQLEAFVDLR	0.29		0.33	13.2
83	Pyruvate kinase I	P14178	LNFSHGDYAEHGQR	0.41		0.33	23.1
84	ABC transporter ATP-binding protein yjjK	P37797	ESIEEAVSEVVNALK	0.32		0.33	3.9
85	Pyruvate kinase II	P21599	GLPADVVPGDILLLDDGR	0.35		0.33	4.5
86	Osmotically inducible lipoprotein E	P23933	AQVAQIAGKPSSEVSMIHAR	0.28		0.33	17.4
87	Polyribonucleotide nucleotidyltransferase	P05055	LHILGVMEQAINAPR	0.39		0.33	17.7
88	50S ribosomal protein L16	P02414	VLYEMDGVPEELAR	0.34		0.33	1.2
89	Aconitate hydratase 1	P25516	VLMQDFTGVPAVVDLAAMR	0.36		0.33	7.8
90	Aspartate aminotransferase	P00509	MFENITAAPADPILGLADLFR	0.28		0.33	16.5
91	Hypothetical protein yfcZ	P76504	ITPTFTEESDGVR	0.39		0.33	18.3
92	30S ribosomal protein S15	P02371	TAQINHLQGHFAEHK	0.35		0.33	5.7
93	ATP synthase delta chain	P00831	LNALPDVLEQFIHLR	0.36		0.33	7.5
94	50S ribosomal protein L29	P02429	SVEELNTELLNLLR	0.34		0.33	1.8
95	Ribose-phosphate pyrophosphokinase	P08330	TLTLSGMLAEAIR	0.31		0.33	8.4
96	Triosephosphate isomerase	P04790	SATPAQAQAVHK	0.36		0.33	8.1
97	Succinyl-CoA synthetase alpha chain	P07459	GGTTHLGLPVFNTVR	0.29		0.33	11.7
98	Periplasmic oligopeptide-binding protein	P23843	AQGNMPAYGYTPPYTDGAK	0.31		0.33	8.1
99	Isocitrate dehydrogenase	P08200	STQVYGQDVWLPAETLDLIR	0.32		0.33	4.8

100	Hypothetical protein yfeA	P23842	AALSVL	0.38	0.33	12.9
101	Xaa-Pro aminopeptidase	P15034	EIYDIVLESLETSLR	0.3	0.33	11.4
102	Preprotein translocase secA subunit	P10408	ALVDEVDSILIDEAR	0.32	0.33	3.6
103	Phosphoenolpyruvate-protein phosphotransferase	P08839	IMFPMIISVEEVR	0.29	0.33	14.1
104	Inosine-5'-monophosphate dehydrogenase	P06981	LNIPMLSAAMDTVTEAR	0.39	0.33	16.8
105	Catalase HPII	P21179	GPTLLEDFILR	0.34	0.33	0.9
106	Hypothetical acetyltransferase yhbS	P45473	VEIPIDAPGIDALLR	0.35	0.33	6.0
107	MTA/SAH nucleosidase	P24247	IGIIGAMEEEVTLLR	0.3	0.33	11.4
108	Cytochrome d ubiquinol oxidase subunit I	P11026	YHFEQSSTTTQPAR	0.4	0.33	18.6
109	Hypothetical protein yccJ	P46131	TDKDSLFWGEQTIER	0.38	0.33	14.4
110	Zinc-type alcohol dehydrogenase-like protein yahK	P75691	ADQINEAYER	0.39	0.33	16.2
111	Phosphopentomutase	P07651	AAGLELFDR	0.27	0.33	19.2
112	50S ribosomal protein L14	P02411	MIQEQTMLNVADNSGAR	0.33	0.33	0.3
113	Hypothetical protein yggE	P11668	TQPDYDYQDGK	0.28	0.33	16.2
114	Dihydrolipoyl dehydrogenase	P00391	YDAVLVAIGR	0.37	0.33	11.4
115	ATP synthase B chain	P00859	AEAEQER	0.37	0.33	11.7
116	HIT-like protein ycfF	P36950	IAEQEGIAEDGYR	0.38	0.33	13.2
117	30S ribosomal protein S13	P02369	FVVEGDLR	0.4	0.33	21.3
118	50S ribosomal protein L13	P02410	DWYVVDATGK	0.29	0.33	13.2
119	Histone-like protein HLP-1	P11457	AQAFEQDR	0.37	0.33	10.2
120	Transketolase 1	P27302	HNPQNPSWADRDR	0.28	0.33	14.7
121	Fructose-bisphosphate aldolase class I	P71295	AINYGYTDDR	0.36	0.33	9.9
122	ADP-L-glycero-D-manno-heptose-6-epimerase	P17963	FLFDEYVR	0.38	0.33	12.6
123	Glucosamine--fructose-6-phosphate aminotransferase	P17169	DVAEILLEGLR	0.36	0.33	6.6

124	50S ribosomal protein L22	P02423	LVADLIR	0.32	0.33	3.3
125	Elongation factor P	P33398	VPLFVQIGEVIK	0.35	0.33	4.8
126	Protein ydhD	P37010	FAYVDILQNPDIR	0.3	0.33	8.7
127	50S ribosomal protein L11	P02409	GLPIPVVITVY	0.26	0.33	21.6
128	30S ribosomal protein S19	P02375	LGEFAPTR	0.3	0.33	9.9
129	30S ribosomal protein S11	P02366	ALNAAGFR	0.36	0.33	9.3
130	30S ribosomal protein S2	P02351	AGVHFGHQTR	0.35	0.33	4.8
131	30S ribosomal protein S20	P02378	AFNEMQPIVDR	0.41	0.33	22.8
132	Hypothetical protein ybeL	P46129	ESLWQELADITDK	0.33	0.33	0
133	Adenylosuccinate synthetase	P12283	TGWLDTVAVR	0.38	0.33	15.3
134	Glutathione S-transferase	P39100	FRPDTPEEYKPTVR	0.29	0.33	14.1
135	Elongation factor Ts	P02997	AQFEEER	0.30	0.33	8.7
136	Major outer membrane lipoprotein	P02937	VDQLSNDVNAMR	0.32	0.33	4.5
137	Protein ygiW	P52083	DDTWVTLR	0.36	0.33	8.1
138	Universal stress protein A	P28242	AVSMARPY	0.39	0.33	16.5
139	50S ribosomal protein L17	P02416	LFNELGPR	0.37	0.33	12
140	10 kDa chaperonin	P05380	MNIRPLHDR	0.38	0.33	14.4
141	Citrate synthase	P00891	LMGFGHR	0.34	0.33	2.7
142	D-ribose-binding periplasmic protein	P02925	ERGEGFQQAVAAHK	0.35	0.33	5.1
143	PTS system, galactitol-specific IIA component	P37187	GVVHDTWPQALIAR	0.39	0.33	16.5
144	50S ribosomal protein L18	P02419	SGFQYHGR	0.4	0.33	19.5
145	Osmotically inducible protein C	P23929	KGQAHWEGDIK	0.39	0.33	19.8
146	Protein hdeA	P26604	QANFKDK	0.38	0.33	12.9
147	Hypothetical protein ygaM	Q47413	GEAEAAR	0.41	0.33	21.9
148	Glucose-6-phosphate isomerase	P11537	EVVEQEYR	0.36	0.33	7.5
149	Universal stress protein F	P37903	TILVPIDISDSELTQR	0.26	0.33	21.3

150	50S ribosomal protein L24	P60624	HQKPVPALNQPGGIVEK	0.34	0.33	0.6
151	Superoxide dismutase [Mn]	P00448	EFWNVVNWDEAAAR	0.36	0.33	9
152	Trigger factor	P22257	ANDIDVPAALIDSEIDVLR	0.29	0.33	14.1
153	Peptidase B	P37095	LMIIDWVR	0.29	0.33	12.3
154	3-methyl-2-oxobutanoate hydroxymethyltransferase	P31057	IEGGEWLVETVQMLTER	0.29	0.33	14.1
155	Inorganic pyrophosphatase	P17288	VEGWENAEAAK	0.35	0.33	3.6
156	50S ribosomal protein L2	P60422	GTAMNPVDHPHGGGEGR	0.31	0.33	5.7
157	ATP-dependent hsl protease ATP-binding subunit hslU	P32168	NNWGQTEQQQEPSAAR	0.39	0.33	16.5
158	Hypothetical protein ybhE	P52697	EGFQPTETQPR	0.39	0.33	16.8
159	30S ribosomal protein S6	P02358	LEDWGR	0.36	0.33	8.7
160	Aspartate-semialdehyde dehydrogenase	P00353	NVGFIGWR	0.34	0.33	3

Please note that the average ratio of all measured pairs is 0.34 and relative standard deviation is 10.3%

* Peptide sequences were confirmed by manually interpreting the overlaid MS/MS spectra of a peptide pair, after a relatively high matching score (>20) was generated from the database search.

** Obs'd: observed abundance ratio from the peptide pair; Exp'd: expected abundance ratio; %SD: relative standard deviation of the observed abundance ratios; %Err: percent error calculated from the average of the observed abundance ratios and the expected ratio.

Chapter 4

Identification and Quantification of Differentially Expressed Proteins in E-Cadherin Deficient SCC9 Cells and SCC9 Transfectants Expressing E-Cadherin by Dimethyl Isotope Labeling, LC-MALDI MS and MS/MS [a]

4.1 Introduction

Due to protein heterogeneity and diversity, proteome investigations have been designed to characterize proteins at different levels, including their identification [1, 2], post-translational modifications (PTMs) [3-5], relative abundance levels [6, 7], and interactions with other proteins [8]. Among them, the most interest has been focused on methods designed to measure global changes in relative protein abundance between two distinct proteome samples. Global profiling of the entire proteome of cells or tissues in different states ideally requires a combination of techniques to achieve both confident identification and accurate quantification of complex protein mixtures in a high-throughput manner. However, the realization of this goal has turned out to be a difficult challenge. This is mainly due to the extremely diverse physicochemical properties and wide dynamic range of proteins expressed in cells or tissues, with the additional complication of PTMs.

---

[a] A portion of this chapter is published as: C. Ji, L. Li, M. Gebre, M. Pasdar and L. Li, "Identification and Quantification of Differentially Expressed Proteins in E-Cadherin Deficient SCC9 Cells and SCC9 Transfectants Expressing E-Cadherin by Dimethyl Isotope Labeling, LC-MALDI MS and MS/MS" *J. Proteome Res.* **2005**, 4, 1419-1426. Dr. Li from Dr. Pasdar' group grew SCC9 and SCC9-E cells and also did immunoblotting and immunoflurescence experiments. Mr. M. Gebre helped in SCX and LC-MALDI experimental part.

Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) [9, 10] has been widely used for studies of protein expression in cells and tissues. This approach has been playing an essential role in proteomic studies. However, 2D PAGE is time-consuming, laborious and difficult for analysis of highly hydrophobic or basic proteins, as well as, proteins with very high or very low molecular weight. In the past few years, several methods based on stable isotopic labeling of proteins or peptides have been reported [11-14]. Among these methods, the isotope-coded affinity tag (ICAT) approach pioneered by Aebersold and coworkers [11, 15-17] has been extensively used. To improve the performance of the isotope labeling approach, alternative labeling reagents for peptides or in- culture metabolic labeling of proteins have been reported for the quantification of global protein expression [13, 18-29].

In Chapter 5, the compatibility of dimethyl labeling of the amino groups in peptides, via reductive amination [18, 30], with liquid chromatography (LC) matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS), using a recently developed heated droplet interface [31], was evaluated. Using standard protein mixtures and *E. coli* extracts of known quantities, it was demonstrated that the approach, based on dimethyl labeling and LC-MALDI MS, had the following advantages. First, the high sample loading capacity of a microbore LC column, compared to conventional approaches of using capillary LC columns [28, 32, 33], gives rise to accurate protein quantification due to the increased signal-to-noise ratio for a protein mixture with a concentration dynamic range of as high as 1 × 10^4. Second, unlike electrospray ionization (ESI)-based methods, the MALDI method separates protein identification from protein quantification, thereby allowing the selection of only the peptides from the differentially expressed proteins to be identified, which saves sample, instrument and data analysis time. Third, differential dimethyl labeling using d(0)- or d(2)-formaldehyde is efficient and relatively inexpensive. The reagents are commercially available.

In this work, successful application of the integrated strategy of dimethyl labeling and LC-MALDI MS for determining the proteins differentially expressed between an E-cadherin-deficient human carcinoma cell line (SCC9) and its transfectants expressing E-cadherin (SCC9-E) is reported. The effective use of this approach as a quantitative tool for protein expression analysis was validated by comparing the MS quantitative results to those obtained by immunoblotting and immunofluorescence assays for six proteins with differential expression. E-cadherin is the epithelial-specific, calcium-dependent cell adhesion molecule, which plays critical roles in development and in maintenance of the integrity and function of epithelial tissues [34]. The cytoplasmic domain of E-cadherin binds to proteins called catenins [35]. β-Catenin and γ-catenin bind in a mutually exclusive manner to the distal portion of the cytoplasmic domain of

E-cadherin and connect it via α-catenin to the actin cytoskeleton [34]. The association of E-cadherin with the actin cytoskeleton is essential for the adhesive function. Recent studies have shown that the disruption of the cadherin-catenin complex can activate cellular signaling pathways that regulate cell proliferation, differentiation, motility and programmed cell death.

The SCC9 cell line is a human carcinoma cell line, which exhibits a transformed morphology and is moderately invasive. Previous study has shown that SCC9 cells are E-cadherin-deficient, do not form cell-cell junctions, have fibroblastoid morphology and lack contact inhibition of growth [36]. Exogenous expression of E-cadherin in this cell line (SCC9-E) led to decreased growth, increased adhesiveness, induction of an epidermoid phenotype and contact inhibition of growth [36]. E-cadherin expression also led to decreased expression of mesenchymal markers and increased expression of a number of epithelial- specific proteins [36]. In this work, the identification of 49 proteins exhibiting steady-state level changes of greater than 2-fold between SCC9 and SCC9-E cells is reported. The functional implications of these findings are discussed.

4.2 Experimental

4.2.1 Chemicals and Reagents

Sodium dodecyl sulfate (SDS), d(0)-formaldehyde (37% (w/w)), sodium cyanoborohydride, bovine trypsin and trifluoroacetic acid (TFA) were purchased from

Sigma-Aldrich (Oakville, ON, Canada). d(2)-Formaldehyde (~20% (w/w) in deuterated water) was obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). HPLC grade acetone and acetonitrile were purchased from Fisher Scientific Canada (Edmonton, AB, Canada). Water used in these experiments was obtained from a Milli-Q Plus purification system (Millipore, Bedford, MA). The MALDI matrix 2,5-dihydroxybenzoic acid (DHB) was purchased from Aldrich (Milwaukee, WI).

4.2.2 Cell culture and sample preparation

The E-cadherin-deficient human squamous carcinoma cell line SCC9 was obtained from the American Type Culture Collection (ATTC) (Rockville, MD). Stable SCC9 transfectants expressing E-cadherin (SCC9-E) have been described previously [36]. Cells were maintained in minimum essential medium (MEM) (Sigma, St. Louis, MO) supplemented with 10% fetal bovine serum (FBS).

4.2.3 Protein extraction

Cells were grown to confluence in 150 mm tissue culture dishes. The cells were washed three times with 150 mM sorbitol in 10 mM phosphate buffer (pH 7.0), scraped with a rubber policeman and centrifuged at 800 ×g for 5 min. The supernatants were discarded and the pellets stored at –80 °C for later use. Proteins were extracted from 10^6 cells using CytoBuster Protein Extraction Reagent (Novagen) according to the manufacturer's directions. The frozen cell pellets were thawed, resuspended in 150 µL

CytoBuster reagent and incubated at room temperature for 5 min. The extracts were centrifuged for 5 min at 15,000 ×g at 4 °C and the supernatants transferred into fresh tubes. Protein concentrations were assayed by BioRad reagent using BSA as the standard.

4.2.4 Acetone precipitation and in-solution protein digestion

The disulfide bonds in 300 μg of protein extracted from the two cell lines, each in a 2 mL siliconized vial were reduced with dithiothreitol (DTT). The free thiol groups were then blocked by reaction with iodoacetamide. The reduced and alkylated extracts were acetone precipitated to remove detergent, unreacted DTT and iodoacetamide. Acetone was pre-cooled at –80 °C and added to the protein extracts to a final concentration of 83% (v/v). After vortexing for 1 min, the mixture was kept at –20 °C for 2 h and then centrifuged for 10 min at 15,000× g at 4 °C. The supernatant was decanted and properly disposed. Acetone was evaporated by incubating the pellet at room temperature. For the in-solution digestion, each pellet was solubilized by the addition of 10 μL 1% SDS (w/v). The mixture was vortexed thoroughly, the SDS diluted to 0.05% and the pH adjusted to ~8.5 with 1 M $NaHCO_3$. Finally, 5 μL of 1 μg/μL bovine trypsin and 10 μL of 20 mM $CaCl_2$ were added to the mixture, and allowed to digest overnight at 37 °C.

4.2.5 Dimethyl labeling of the protein digests

Dimethyl labeling of N-termini and ε-amino groups of lysine residues of the peptides was performed as previously described [37]. Trypsin digestion was stopped by adjusting the pH to 5 with 1% TFA. Then an equal volume of sodium acetate buffer (0.2 M, pH 6.0) was added to each vial and mixed by vortexing followed by the addition of 20 µL of freshly prepared 1 M sodium cyanoborohydride solution. After vortexing, the SCC9 and SCC9-E protein digests were labeled with 10 µL of 4% (in water) d(0)- or d(2)-formaldehyde, respectively. The labeled mixtures were vortexed and incubated at 37 °C for 2 h. When necessary, excess formaldehyde was consumed by addition of ammonium bicarbonate. After labeling, the solution in each vial was acidified to pH 3 by adding 1% TFA. Following acidification, the d(0)-formaldehyde-labeled SCC9 cells protein digests were pooled with d(2)-formaldehyde-labeled protein digests of SCC9-E cells. The mixtures were stored at –80 °C for further analysis. Caution: sodium cyanoborohydride is a highly toxic compound that releases hydrogen cyanide gas upon exposure to strong acid. Therefore, labeling process must be performed in a fume hood.

4.2.6 Cation exchange chromatography

The combined peptide mixtures were separated by strong cation exchange (SCX) chromatography on an Agilent 1100 HPLC system (Palo Alto, CA) using a 2.1 x 150 mm Hydrocell^{TM} SP 1500 column (5 mm, Catalog No.: 24-34 SP, BioChrom Labs, Inc., Terre

Haute, IN). Sample mixtures were diluted 10 fold with SCX mobile phase A (20% v/v acetonitrile in 0.1% TFA). About 1.5 mL proteins digest (~500 μg) was loaded and washed isocratically for 15 min at 0.2 mL/min to remove excess reagents. Peptides were eluted with a linear gradient of 0-30% B (20% v/v acetonitrile in 0.1% TFA, 1 M NaCl) in 7 min and 30-50% B in 2 min at 0.2 ml/min, with fractions collected at 1-min time intervals.

4.2.7 Microbore LC-MALDI QqTOF

Peptides in each SCX fraction (100 μL) were further separated by reversed-phase chromatography on a 1.0 × 150 mm Vydac C_{18} column (5 μm particles with 300 Å pore size, Catalog No.: 218TP5115) at a flow-rate of 40 μL/min in an Agilent 1100 capillary HPLC equipped with an auto-sampler. Gradient elution was performed with solvent A (Milli-Q water, 0.1% TFA and 4% acetonitrile in water, v/v/v) and B (0.1% TFA in acetonitrile, v/v). The gradient was 0-5% B in 5 min, 5-40% B in 85 min, 40-90% B in 15 min. Before gradient elution, peptides in each SCX fraction were desalted by pumping with 100% solution A for 10 min. Immediately after desalting, the HPLC fractions were directly collected in 1-min time intervals onto a 100-well MALDI plate (Applied Biosystems, Concord, ON, Canada) using a home-built, heated droplet LC-MALDI interface [31]. After fractionation, the dried peptides in each well were redissolved and mixed with a DHB matrix (in 50%ACN/50%water) by the addition of 0.8 μL of 100 mg/mL. Subsequent MALDI MS and MALDI MSMS data were acquired on an Applied Biosystems/MDS-Sciex QSTAR Pulsar QqTOF instrument equipped with an orthogonal MALDI source employing a 337 nm nitrogen laser (Concord, ON, Canada) that has been previously described [37, 38]. The instrument was operated in positive ion mode. The MS spectrum for each spot on the MALDI plate was collected in an automated mode and collision-induced dissociation (CID) of peptides was achieved with argon as collision gas. Spectra were acquired and processed using Sciex supporting software and re-processed with Igor Pro software (WaveMetrics, Lake Oswego, Oregon, USA).

4.2.8 Relative quantification of peptide pairs

The peak ratios of the peptide pairs were obtained by calculating the ratios of the monoisotopic signal intensities of d(0)- and d(2)-formaldehyde labeled peptide pairs. In this study, MS/MS analysis was done on the peptide pairs with a relative peak intensity difference of greater than 2-fold. The selection of a 2-fold change criterion was based on the quantitative studies of standard protein samples with replicate experiments using the dimethyl labeling LC-MALDI approach [37].

4.2.9 Protein identification from MS/MS data

The MS/MS spectra of both d(0)- and d(2)-formaldehyde labeled peptides were subjected to a proteome database search for protein identification using the Mascot search program (http://www.matrixscience.com). Database searching was restricted to Homo sapiens in Swiss-Prot database. S-acetamido, N-terminal and lysine modifications were selected as fixed, methionine oxidation as variable, 4 missed cleavages allowed and mass tolerance of the precursor peptide and its fragments at ± 0.3 Da. In all cases, MS/MS data were searched twice; in one case with no requirement that the peptides be tryptic, the other being constrained to only tryptic peptides. The search results were confirmed by comparing the fragment ion peak matches using information derived from the isotope tag at the N-termini of the peptides (see Results).

4.2.10 Electrophoresis and Western blotting

Confluent cultures of SCC9 and SCC9-E cells were extracted as described above. Equal amounts of total protein from each cell line were resolved by SDS-PAGE, transferred to nitrocellulose membranes and processed for immunoblotting. Membranes were incubated with antibodies at the following concentrations: α-catenin (Sigma), $p120^{ctn}$ (Sigma), colligin (Stressgen, Canada) and peroxiredoxin (Biomol, USA), 1:1000, caldesmon (BD Biosciences, USA), 1:2000 and calreticulin (Dr. Marek Michalak, University of Alberta, Canada), 1:3000, tubulin (Developmental Studies Hybridoma Bank, NCI, USA), 1:300. Primary antibodies were detected by species-specific horseradish peroxidase (HRP)-conjugated secondary antibodies (Sigma) at 1:2000 (colligin, caldesmon, peroxiredoxin and tubulin) or 1:3000 (calreticulin) or 1:5000 (α-catenin, $p120^{ctn}$) dilution and developed by ECL (Amersham, USA). Blots were scanned and each protein band quantified using NIH imager software. Histograms were constructed by normalizing the value obtained for each protein in each cell line to the amount of tubulin detected in the same line relative to that of the SCC9 cells. Each experiment was repeated 3 times and the results from one typical experiment are presented.

4.2.11 Immunofluorescence

Cells were plated on glass coverslips and grown to confluence, fixed with 3.7% (v/v) formaldehyde for 15 min, permeabilized with 0.25% (v/v) Triton X-100 for 2 min on ice and blocked for 1 h with 4% (v/v) goat serum and 50mM NH_4Cl in PBS. Replicate coverslips were incubated with α-catenin or $p120^{ctn}$ (1:200), caldesmon (1:100), calreticulin (1:75), peroxiredoxin (1:200) and colligin (1:100). All dilutions were made in PBS. Coverslips were incubated with various antibodies for 1 h followed by fluorochrome-conjugated, species-specific secondary antibodies for 20 min. Coverslips were mounted in elvenol containing 0.2% (w/v) paraphenylene diamine (PPD, pH 8.0) and viewed with a 100x objective using an Olympus BX50 fluorescence microscope. Images were captured with a digital fluorescence camera (Spot Diagnostic instruments).

4.3 Results and Discussion

Proteins extracted from SCC9 and SCC9-E cells were tryptically digested, labeled with either d(0)- or d(2)-formaldehyde. The combined peptide mixture was separated by a narrowbore SCX column. Figure 4.1A shows the UV absorbance chromatogram of the dimethyl labeled tryptic peptides from the ion exchange chromatographic separation. Peptides were separated into seven 1-min fractions with collection starting at 17 min. Each SCX fraction was further separated by a microbore RP-LC column, with direct spotting of 1 min eluate onto a 100-well MALDI target using a home-built heated droplet LC-MALDI interface. Figure 4.1B shows the UV absorbance chromatogram of the RP-LC separation of the SCX fraction collected between 19 and 20 min. After the addition of MALDI matrix (DHB), each spot on the MALDI target was initially analyzed by MALDI QqTOF MS with a MS scan. After the MS scan, the ratios for all peptides were determined as previously described [37]. For those peptide pairs exhibiting larger mass differences, the ratios were determined by calculating the ratios of monoisotopic peak intensities of d(0)- and d(2)-formaldehyde labeled peptide pairs. For larger peptide pairs with small mass differences, which result in significant overlap of the isotope envelopes, the overlapped peptide pairs were de-convoluted by the MS-isotope program. This is an online tool for calculating and visualizing isotope patterns of peptides provided by the UCSF Mass Spectrometry Facility (http://prospector.ucsf.edu/ucsfhtml4.0/msiso.htm). The ratios of peptide pairs were determined using de-convoluted monoisotopic peak intensities of d(0)- and d(2)-formaldehyde labeled peptide pairs.

Figure 4.1C shows the MALDI mass spectrum of one RP-LC fraction at 28 min as characterized by the MALDI QqTOF mass spectrometer. Peak pairs separated by

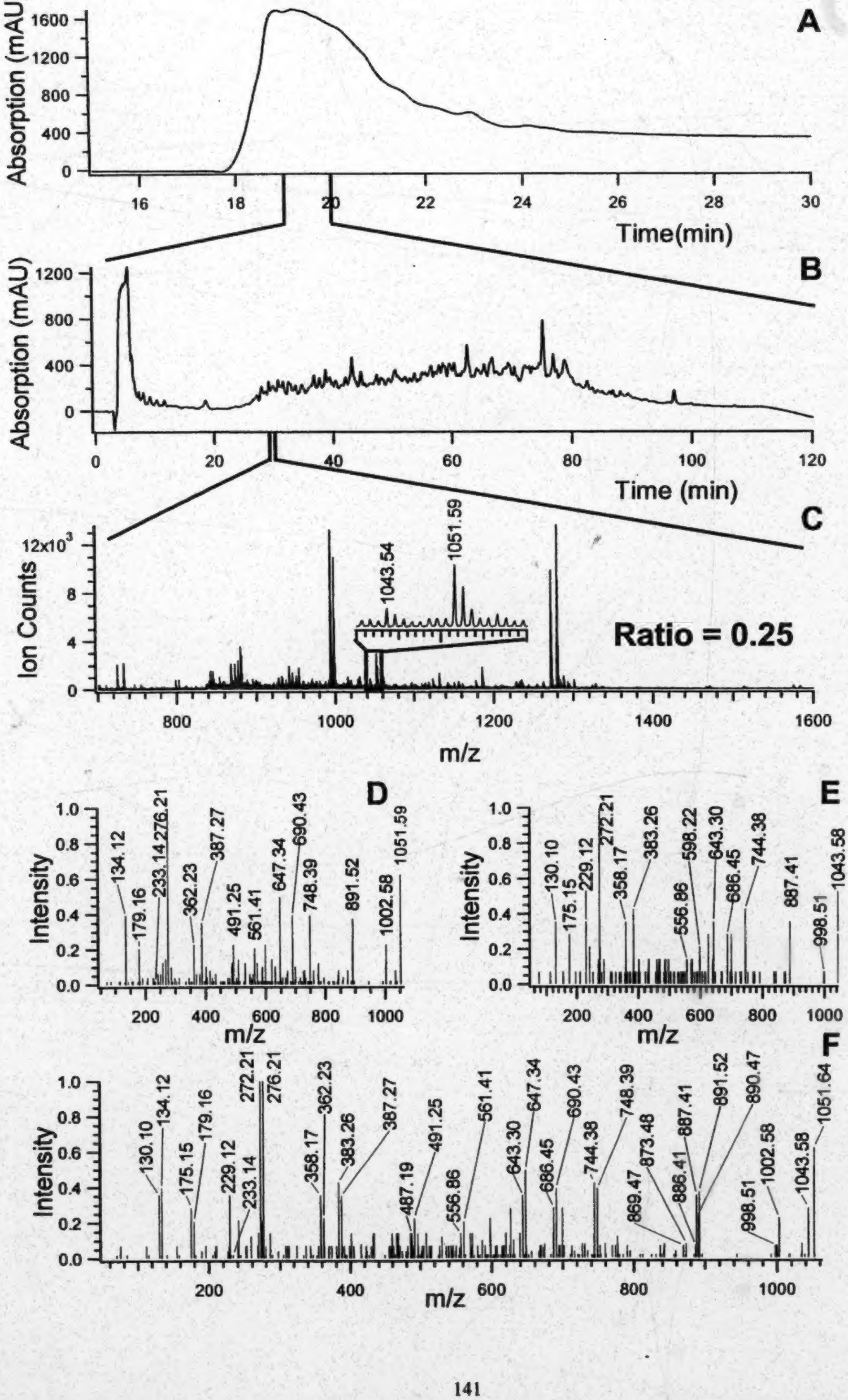
A
Absorption (mAU)
0
400
800
1200
1600
16
18
20
22
24
26
28
30
Time(min)
B
Absorption (mAU)
0
400
800
1200
0
20
40
60
80
100
120
Time (min)
C
Ion Counts
12x10^3
8
4
0
1043.54
1051.59
Ratio = 0.25
800
1000
1200
1400
1600
m/z
D
Intensity
0.0
0.2
0.4
0.6
0.8
1.0
134.12
179.16
233.14
276.21
362.23
387.27
491.25
561.41
647.34
690.43
748.39
891.52
1002.58
1051.59
200
400
600
800
1000
m/z
E
Intensity
0.0
0.2
0.4
0.6
0.8
1.0
130.10
175.15
229.12
272.21
358.17
383.26
556.86
598.22
643.30
686.45
744.38
887.41
998.51
1043.58
200
400
600
800
1000
m/z
F
Intensity
0.0
0.2
0.4
0.6
0.8
1.0
130.10
134.12
175.15
179.16
229.12
233.14
272.21
276.21
358.17
362.23
383.26
387.27
487.19
491.25
556.86
561.41
643.30
647.34
686.45
690.43
744.38
748.39
869.47
873.48
886.41
887.41
891.52
890.47
998.51
1002.58
1043.58
1051.64
200
400
600
800
1000
m/z

Figure 4.1 Two-dimensional LC MALDI MS and MS/MS analysis of differentially expressed proteins between SCC9 and SCC9-E cells. (A) SCX chromatogram of d(0)- and d(2)-formaldehyde labeled peptides mixture. UV absorbance was recorded at 214 nm. (B) RP chromatogram of d(0)- and d(2)-formaldehyde labeled peptides in one SCX fraction collected between 19 and 20 min as indicated in (A). UV absorbance was recorded at 214 nm. (C) MALDI mass spectrum obtained for the fraction collected at 28 min. The inset shows the expanded view of the peptide pair as indicated by the rectangular. (D) and (E) MALDI MS/MS spectra for a pair of peptides at m/z 1043.54 and 1051.59. (F) Overlaid MS/MS spectra of (D) and (E).

mass differences of $\Delta = 4.024n$ in the MS spectrum were selected, where n is the number of dimethyl labeling site(s) in the peptide sequence (one for the N terminus and one for each lysine in the peptide). From this MALDI MS spectrum, 23 peptide pairs were quantitatively analyzed. However, only two peptide pairs within the abundance ratio appeared to result from differentially expressed proteins (i.e., their relative abundance changes were greater than 2-fold). Therefore, only these two peptide pairs were selected for CID or MS/MS to identify the differentially expressed proteins. Sequence database searching was subsequently performed for each peptide pair to identify the corresponding proteins. Using this abundance ratio-dependent approach, most of the effort including instrumental and data analysis time was spent on identifying biologically interesting proteins, i.e., the differentially expressed proteins between the two distinct samples. In this study, 5480 peptide pairs were detected in the MALDI MS scan. However, only 320 pairs showed relative abundance ratios of greater than 2. They were selected for CID, followed by MS/MS database searching to identify those proteins expressed differentially between SCC9 cells and its transfectants (SCC9-E).

Two representative MS/MS spectra of a peptide pair [d(0)-EAEEREPK and d(2)-EAEEREPK] are shown in Figures 4.1D and E. An example of using the labeled peptide pairs to confirm peptide sequence results generated from database search is shown in Figure 4.1F. In this case, the mass differences between all the fragment ions generated from the N-terminus or C-terminus are uniformly 4.024 Da, which can be used as a mass tag to confirm the identified peptide sequence. This is one of the advantages of dimethyl labeling where the labeled amine groups provide a mass tag that assists in MS/MS spectral interpretation [37].

Using this strategy, 49 differentially expressed proteins were identified when the protein profile of SCC9 cells was compared to that of SCC9-E cells. Table 4.1 lists these proteins along with some of their properties. I believe that the strategy described in this study has the potential to provide more comprehensive information than ICAT and 2-DGE, the two most popular techniques currently used for this type of analysis. For example, four differentially expressed proteins, namely putative RNA-binding protein 3, neddylin, prothymosin alpha and keratin, which do not have any cysteine residues and would have been missed by the ICAT method, have been quantified and identified using the approach described herein. In addition, some proteins with low molecular weight (e.g., neddylin, 9.3 kDa) were identified, which would not have been easily identified by 2-DGE analyses.

As Table 4.1 shows, E-cadherin expression in SCC9 cells lead to the downregulation of 8 and upregulation of 41 proteins. Half of the downregulated proteins, namely DNA replication licensing factor MCM7, deoxycytidine kinase, nucleosome assembly protein 1-like 1 (NAP-1 related protein) (hNRP) and prothymosin alpha, are directly involved with cell proliferation [39-42] and have been shown to be upregulated in various tumors or carcinoma cell lines [43-48]. The remaining downregulated proteins are involved in actin polymerization, protein folding or metabolic

Table 4.1 List of identified proteins differentially expressed between SCC9 and SCC9-E cells.

#	Access ID	Ratio* d(2):d(0)	Peptides identified**	Protein name	Mass (kDa)	Subcellular location
Downregulated						
1	Q9UHB6	0.37	GNYDEGFGHRPHK SKGNYDEGFGHRPHK	Epithelial protein lost in neoplasm	87.7	Cytoplasmic
2	P32119	0.31	KEGGLGPLNIPLLADVTR	Peroxiredoxin 2	22.5	Cytoplasmic
3	P33993	0.40	LAQHITYVHQHSR	DNA replication licensing factor MCM7	82.8	Nuclear (By similarity).
4	P50990	0.47	ELEVQHPAAK	T-complex protein 1, theta subunit	61.3	Cytoplasmic
5	P06454	0.46	AAEDDEDDDVDTKK	Prothymosin alpha	12.3	Nuclear
6	P27707	0.34	HESWLLHR	Deoxycytidine kinase	31.4	Nuclear
7	P55209	0.43	AKIEDEKKDEEKEDPK	Nucleosome assembly protein 1-like 1	46.7	Nuclear
8	O60488	0.37	AKPTSDKPGSPYR	Long-chain-fatty-acid--CoA ligase 4	82	Endomembrane
Upregulated						
1	P07355	2.04 ±0.15	SVPHLQK AYTNFDAER QDIAFAYQR SALSGHLETVILGLLK LSLEGDHSTPPSAYGSVK AYTNFDAERDALNIETAIK	Annexin A2	39.6	In the lamina beneath the plasma membrane
2	P04406	2.89 ±0.25	FHGTVK LTGMAFR VVDLMAHMASKE LISWYDNEFGYSNR VIISAPSADAPMFVMGVNHEK	Glyceraldehyde-3-phosphate dehydrogenase, liver	36.8	Cytoplasmic
3	P08195	2.74 ±0.19	HWDQNER WWHTGALYR IKVAEDEAEAAAAAK IGDLQAFQGHGAGNLAGLK VKDALEFWLQAGVDGFQVR SLLHGDFHAFSAGPGLFSYIR	4F2 cell-surface antigen heavy chain	58.8	Membrane

4	P50454	6.10 ±0.56	GVVEVTHDLQK LQIVEMPLAHK LFYADHPFIFLVR TGLYNYYDDEKEK LSSLIILMPHHVEPLER QHYNCEHSK KPAAAAAPGTAEK TDGALLVNAMFFKPHWDEK AVLSAEQLRDEEVHAGLGELLR	Collagen-binding protein 2 precursor	47.5	ER
5	P00354	3.51 ±0.29	FHGTVK LTGMAFR VGVDGFGR VVDLMAHMASKE LEKPAKYDDIKK	Glyceraldehyde-3-phosphate dehydrogenase, muscle	36.8	Cytoplasmic
6	Q9P2E9	3.84	KAEGTPNQGK KVEGAQNQGK	Ribosome-binding protein 1	157.9	ER (By similarity)
7	Q00610	2.02	LEKHELIEFR RKDPELWGSVLLESNPYR	Clathrin heavy chain 1	196	Cytoplasmic face of coated pits and vesicles
8	P02786	2.16	HVFWGSGSHTLPALLENLK EAGSQKDENLALYVENQFR	Transferrin receptor protein 1	86.7	Membrane
9	P98179	2.16	YYDSRPGGYGYGYGR GFGFITFTNPEHASVAMR	Putative RNA-binding protein 3	17.3	Cytoplasmic
10	Q02809	3.08	LTHYHEGLPTTR HTLGHLLSLDSYR	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1 precursor	85.1	ER
11	Q05682	3.77	EAEGAPQVEAGKR QKEFDPTITDASLSLPSR	Caldesmon	96.2	Cytoplasmic
12	O15460	9.18	KGTAVFWYNLLR SQVLDYLSYAVFQLGDLHR	Prolyl 4-hydroxylase alpha-2 subunit precursor	62.1	ER
13	P02452	>10***	GDKGETGEQGDR YHDRDVWKPEPCR GDRGETGPAGPPGAPGAPGAPG-PVGPAGK	Collagen alpha 1(I) chain precursor	141.7	
14	Q16790	>10	QLHTLSDTLWGPGDSR FPAEIHVVHLSTAFAR	Carbonic anhydrase IX precursor	50.2	Found on the surface microvilli and in the nucleus, particularly in nucleolus
15	P06748	2.55	MTDQEAIQDLWQWR	Nucleophosmin	33.7	Nuclear

		±0.21	DSKPSSTPR FINYVK			
16	Q01995	>10	HVIGLQMGSNR	Transgelin	23.1	Cytoplasmic (Probable).
17	O60716	2.03	GIPVLVGLLDHPK	p120 catenin	110.2	Cytoplasmic and nuclear in particular cells
18	Q15843	2.03	ILGGSVLHLVLALR	Neddylin	9.3	
19	P60174	2.13	TATPQQAQEVHEK	Triosephosphate isomerase	27.4	
20	P27797	2.04	KVHVIFNYK KIKDPDASKPEDWDER	Calreticulin precursor	49.5	ER
21	Q13428	2.23	ELLPLIYHHLLR	Treacle protein	148.9	Nuclear; nucleolar (Potential)
22	Q01650	2.95	ALAAPAAEEKEEAR	Large neutral amino acids transporter small subunit 1	56.3	Integral membrane protein (Probable)
23	Q13509	2.22	LHFFMPGFAPLTAR	Tubulin beta-4 chain	51.3	Cytoplasmic
24	Q03135	2.08	YVDSEGHLYTVPIR	Caveolin-1	20.9	Membrane
25	Q9NZN4	2.52	LLEALDEMLTHDIAK	EH-domain containing protein 2	62.3	
26	P35221	3.50	EKQDETQTK	Alpha-1 catenin	102.8	Peripheral membrane
27	Q8N8S7	3.09	VHIYHHTGNNTFR	Enabled protein homolog	67.3	
28	Q9Y4P3	2.56	EKPQQHNFTHR	Transducin beta-like 2 protein	51.3	
29	Q92896	4.03	EAEEREPK	Golgi apparatus protein 1 precursor	140.8	Golgi
30	Q86UP2	4.41	TAEHEAAQQDLQSK	Kinectin	162.2	Endomembrane
31	Q13162	2.15	IPLLSDLTHQISK	Peroxiredoxin 4	30.5	Cytoplasmic
32	P13674	3.69	FILAPAKQEDEWDKPR	Prolyl 4-hydroxylase alpha-1 subunit precursor	62.7	ER
33	P20908	>10	GVQGPPGPAGKPGR	Collagen alpha 1(V) chain precursor	187.4	
34	O00469	4.06	VVFAADGILWPDKR	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 precursor	84.7	ER
35	P02461	5.85	KHWWTDSSAEKK	Collagen alpha 1(III) chain precursor	141.7	
36	P08133	>10	GLGTDEDTIIDIITHR	Annexin A6	77.7	Stress fibers (By similarity)
37	Q14247	2.75	HESQQDYSK	Src substrate cortactin	63.2	Cytoplasmic

38	P61619	2.05	CVILPEIQKPER	Protein transport protein Sec61 alpha subunit isoform 1	53.1	ER
39	P09493	2.31	AEQAEADKK	Tropomyosin 1 alpha chain	33.8	Cytoplasmic
	P07951			Tropomyosin beta chain	34.1	Cytoplasmic
40	P17252	2.16	EHAFFR	Protein kinase C, alpha type	79.5	Cytoplasmic
	P05771			Protein kinase C, beta type	79.6	Cytoplasmic and membrane-associated (By similarity)
41	P05783	3.45	IREHLEK	Keratin, type I cytoskeletal 18	47.9	Cytoplasmic

* For the proteins quantified based on more than one peptide pair, the average abundance ratio determined from all peptide pairs is listed.

** Peptide sequences were confirmed by manually interpreting the overlaid MS/MS spectra of a peptide pair after a relatively high matching score (>20) was initially generated from the MS/MS database search.

*** The peak intensity of one peptide in a peptide pair was off-scale.

and oxidative pathways, which exhibit increased expression in tumors and transformed cells [49-52]. The reduced expression of these proteins is consistent with decreased growth and the induction of an epidermoid morphology in SCC9 cells expressing E-cadherin [36].

In agreement with the more differentiated, epidermoid morphology of SCC9E cells, a number of proteins with known functions in cell-cell and cell-matrix adhesion, cytoskeletal reorganization, and oxidative state were upregulated (Table 4.1). The upregulated proteins include several collagen or collagen-binding precursor proteins involved in cell-matrix adhesion. Concurrent with the increased collegen levels, colligin (HSP47), a chaperone in the biosynthetic pathway of collagen, was also upregulated [53]. Antibodies to colligin have been shown to increase migration and invasion in SCC cell lines [54]. Upregulated proteins also included a number of cytoskeletal and actin binding proteins such as cortactin, transgelin and caldesmon. Cortactin is an actin binding protein and an important regulator of actin dynamics and cell motility. It also plays an essential role in E-cadherin-actin association and the formation of stable adhesion [55]. Transgelin is a transformation and shape-change sensitive form of the p21 actin binding protein which binds F-actin and causes actin gelation [56]. Ras induced breast and colon carcinomas show decreased levels of transgelin [57]. Caldesmon stabilizes actin microfilaments and plays important roles in cell shape changes and motility [58].

Interestingly, a number of upregulated proteins are directly involved in E-cadherin transport, localization, stability and function including calreticulin, annexin II, caveolin-1, α- catenin and p120 catenin [34, 59-63]. The upregulated proteins also included those functioning in the oxidative pathways such as peroxiredoxin (Prx) IV. Prx IV is a member of the antioxidant family of enzymes and functions by activating intracellular pathways involved in apoptosis [64], consistent with decreased growth rate and increased apoptosis in SCC9-E cells [36].

The differences in the expression of 6 proteins between the two cell lines were validated by immunoblotting (Figure 4.2) and immunofluorescence microscopy (Figure 4.3). These proteins were selected based on their role in cell adhesion, cytoskeleton reorganization and oxidative processes. In Figure 4.2, equal amounts of total cellular proteins from SCC9 and SCC9-E cells were separated by SDS-PAGE and processed for immunoblotting with antibodies against p120^{ctn}, α-catenin, L-caldesmon, calreticulin, colligin and peroxiredoxin IV. Consistent with the MS results increased steady state levels of all proteins were found. The increased levels of various proteins were also evident by the immunofluorescence staining of the cells from each cell line with the corresponding antibodies (Figure 4.3). In addition, the staining showed significant reorganization and peripheral relocation of the catenins and caldesmon, further supporting the development of epidermoid phenotype upon E-cadherin expression. Consistent with increased cell-matrix adhesion, staining for colligin, a chaperone in collagen biosynthetic pathway, became more intense.

Numerous *in vivo* and *in vitro* studies have shown decreased cellular adhesion and

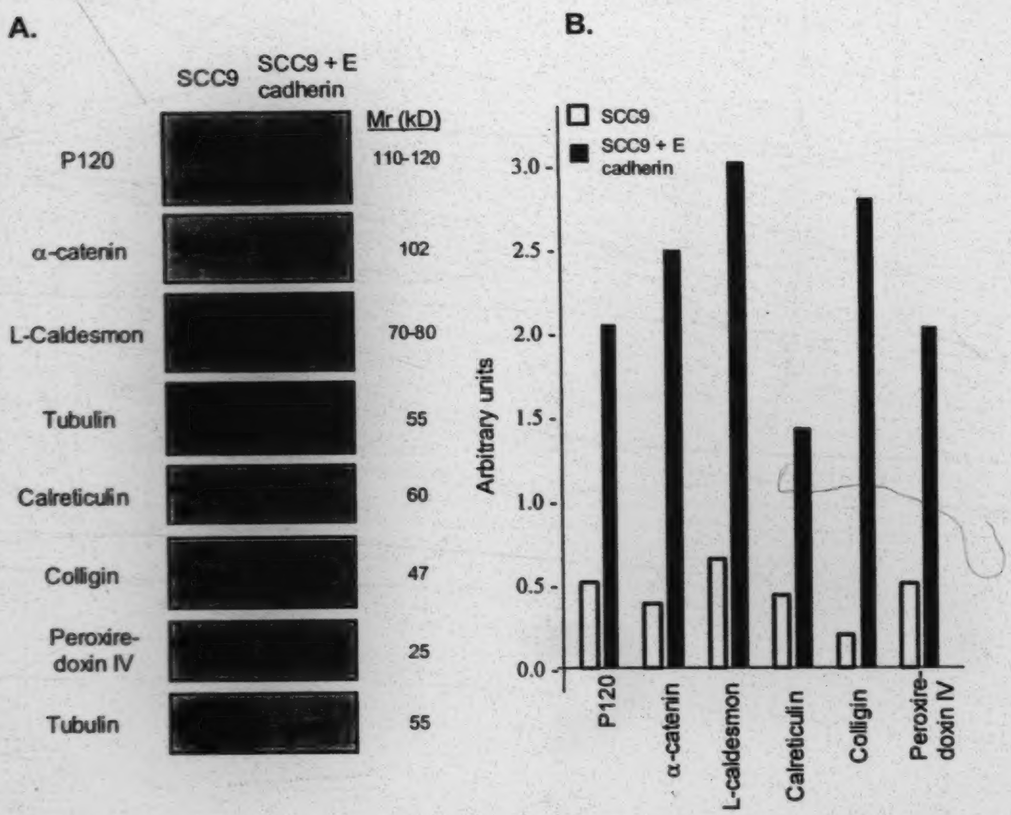

Figure 4.2 Validation of differential expression of selected proteins in E-cadherin-expressing SCC9 cells. (A) Equal amounts of total cellular proteins from SCC9 and SCC9-E cells were resolved on 6% (p120^{ctn}, α-cat, caldesmon and tubulin) or 10% (calreticulin, colligin, peroxiredoxin and tubulin) SDS-gels and processed for immunoblotting with various antibodies as described in Materials and Methods. To confirm equal amounts of protein the same blot processed with tubulin antibodies. (B) Gels from A were scanned and quantified using NIH Image analysis software. Histograms were constructed by normalizing the amount of each protein band to the tubulin in the same cell line relative to the SCC9 cells.

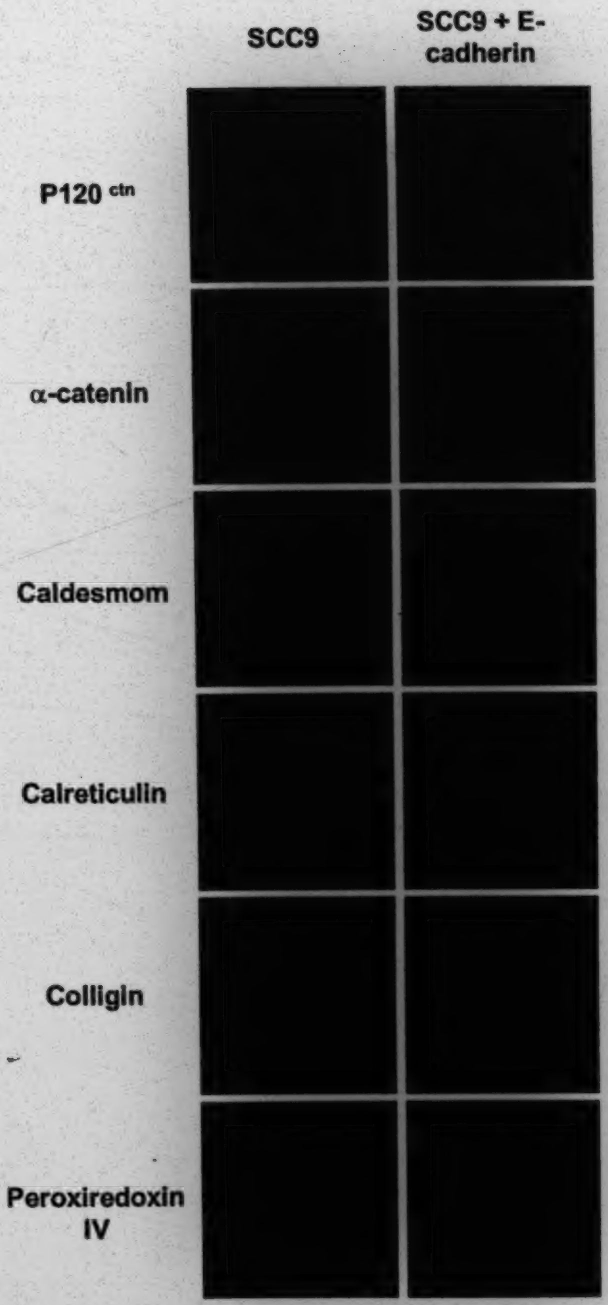
SCC9
SCC9 + E-cadherin
P120 ctn
α-catenin
Caldesmom
Calreticulin
Colligin
Peroxiredoxin IV

Figure 4.3 Subcellular distribution of selected differentially expressed proteins in E-cadherin-expressing SCC9 cells. SCC9 and SCC9-E cells were grown on glass coverslips and processed for immunofluorescence microscopy with p120^{ctn}, α-catenin, caldesmon, calreticulin, colligin or peroxiredoxin antibodies as described in Materials and Methods.

loss of functional E-cadherin in various carcinomas [65, 66], suggesting that disruption of the cadherin-catenin complex may play a causative role in the development and progression of cancer. Cancer proteomics is a powerful approach which allows us to address this question by profiling and identifying proteins that are either differentially expressed or specific to tumors or cancer cell lines [67, 68]. To begin deciphering the molecular basis of the cadherin-mediated signaling in cancer development, the protein profiles of E-cadherin-deficient SCC9 cells and SCC9 transfectants expressing E-cadherin were compared. Since the two lines are genetically identical, the differences in their protein profiles result from the expression of E-cadherin.

Consistent with the development of epidermoid phenotype in SCC9-E cells [36], a number of proteins involved in cell proliferation were downregulated whereas several proteins known to function in cytoskeletal reorganization, cell adhesion and oxidative state of the cell were upregulated. Differential expression of 6 proteins was validated. α-Catenin and p120^{ctn} are known cytoplasmic cofactors of E-cadherin with well-established roles in connecting E-cadherin to the actin network and regulating cell adhesive strength respectively. Increased levels of these proteins were correlated with their peripheral redistribution, as seen in normal epidermal cells [69] in agreement with their roles in mediating normal epidermoid phenotype. Calreticulin decreases motility, increases cell-cell and cell substratum adhesion and spreading [44]. Caldesmon plays important roles in cell shape changes and motility [58], colligin has been implicated in regulation of cell-matrix adhesion [53] and Prx IV is involved in maintaining the cell number by activating intracellular pathways involved in apoptosis [64] and has been shown to be downregulated in gastric adenocarcinomas [70].

In summary, the results clearly show the validity of the differential dimethyl labeling and 2-D LC in conjunction with MALDI QqTOF MS for accurate quantification and identification of differentially expressed protein. In this work, a total of 5480 peptide pairs in a mixture of differentially labeled digests of whole cell lysates from E-cadherin-deficient human carcinoma cell line (SCC9) and its transfectants expressing E-cadherin were examined. Among them, 320 peptide pairs showed relative peak intensity changes of greater than 2-fold. These pairs were selected for MS/MS analysis which led to the identification of 49 differentially expressed proteins between the two cell lines. Assuming the same protein identification rate, i.e., 320 peptide pairs for 49 proteins or 6.53 pairs/protein, could be applied to the 5480 peptide pairs examined, I would have detected and compared about 839 (i.e., 5480/6.53) proteins from the two cell lines. These results suggest that differential dimethyl labeling combined with 2D-LC MALDI MS/MS can be used as an effective means of comparing relative expression changes of a large number of proteins from two proteome samples. Furthermore, the results obtained with the E-cadherin-expressing SCC9 cell line show the feasibility of this model system for characterization of the role of cadherin mediated pathways in regulating growth and integrity of epithelial tissues.

4.4 Literature Cited

(1) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat. Biotechnol.* **2001**, *19*, 242-247.

(2) Garin, J.; Diez, R.; Kieffer, S.; Dermine, J. F.; Duclos, S.; Gagnon, E.; Sadoul, R.; Rondeau, C.; Desjardins, M. *J. Cell Biol.* **2001**, *152*, 165-180.

(3) Gerner, C.; Frohwein, U.; Gotzmann, J.; Bayer, E.; Gelbmann, D.; Bursch, W.; Schulte-Hermann, R. *J. Biol. Chem.* **2000**, *275*, 39018-39026.

(4) Goshe, M. B.; Conrads, T. P.; Panisko, E. A.; Angell, N. H.; Veenstra, T. D.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 2578-2586.

(5) Hanisch, F. G.; Jovanovic, M.; Peter-Katalinic, J. *Anal. Biochem.* **2001**, *290*, 47-59.

(6) Moller, A.; Soldan, M.; Volker, U.; Maser, E. *Toxicology* **2001**, *160*, 129-138.

(7) Monahan, I. M.; Betts, J.; Banerjee, D. K.; Butcher, P. D. *Microbiology* **2001**, *147*, 459-471.

(8) Rain, J. C.; Selig, L.; De Reuse, H.; Battaglia, V.; Reverdy, C.; Simon, S.; Lenzen, G.; Petel, F.; Wojcik, J.; Schachter, V.; Chemama, Y.; Labigne, A.; Legrain, P. *Nature* **2001**, *409*, 211-215.

(9) Klose, J. *Humangenetik* **1975**, *26*, 231-243.

(10) O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007-4021.

(11) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat.*

Biotechnol. **1999**, *17*, 994-999.

(12) Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M.; Regnier, F. *J. Chromatogr. B Biomed. Sci. Appl.* **2000**, *745*, 197-210.

(13) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836-2842.

(14) Ong, S. E.; Kratchmarova, I.; Mann, M. *J. Proteome Res.* **2003**, *2*, 173-181.

(15) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946-951.

(16) Griffin, T. J.; Gygi, S. P.; Ideker, T.; Rist, B.; Eng, J.; Hood, L.; Aebersold, R. *Mol. Cell. Proteomics* **2002**, *1*, 323-333.

(17) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, *20*, 512-515.

(18) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Anal. Chem.* **2003**, *75*, 6843-6852.

(19) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell. Proteomics* **2002**, *1*, 376-386.

(20) Everley, P. A.; Krijgsveld, J.; Zetter, B. R.; Gygi, S. P. *Mol. Cell. Proteomics* **2004**, *3*, 729-735.

(21) Che, F. Y.; Fricker, L. D. *Anal. Chem.* **2002**, *74*, 3190-3198.

(22) Krijgsveld, J.; Ketting, R. F.; Mahmoudi, T.; Johansen, J.; Artal-Sanz, M.; Verrijzer, C. P.; Plasterk, R. H.; Heck, A. J. *Nat. Biotechnol.* **2003**, *21*, 927-931.

(23) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., 3rd *Anal.*

Chem. **2002**, *74*, 1650-1657.

(24) Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. *Anal. Chem.* **2003**, *75*, 1316-1324.

(25) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163-170.

(26) Staes, A.; Demol, H.; Van Damme, J.; Martens, L.; Vandekerckhove, J.; Gevaert, K. *J. Proteome Res.* **2004**, *3*, 786-791.

(27) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U S A* **1999**, *96*, 6591-6596.

(28) Zappacosta, F.; Annan, R. S. *Anal Chem* **2004**, *76*, 6618-6627.

(29) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513-523.

(30) Hermanson, G. T. *Bioconjugate Techniques*; Academic Press: San Diego, CA, 1996.

(31) Zhang, B.; McDonald, C.; Li, L. *Anal. Chem.* **2004**, *76*, 992-1001.

(32) Bodnar, W. M.; Blackburn, R. K.; Krise, J. M.; Moseley, M. A. *J Am Soc Mass Spectrom* **2003**, *14*, 971-979.

(33) Zhen, Y.; Xu, N.; Richardson, B.; Becklin, R.; Savage, J. R.; Blake, K.; Peltier, J. M. *J Am Soc Mass Spectrom* **2004**, *15*, 803-822.

(34) Gooding, J. M.; Yap, K. L.; Ikura, M. *Bioessays* **2004**, *26*, 497-511.

(35) Hatzfeld, M. *Int. Rev. Cytol.* **1999**, *186*, 179-224.

(36) Li, Z.; Gallin, W. J.; Lauzon, G.; Pasdar, M. *J. Cell Sci.* **1998**, *111 (Pt 7)*, 1005-1019.

(37) Ji, C.; Li, L. *J. Proteome Res.* **2005**.

(38) Quach, T. T.; Li, N.; Richards, D. P.; Zheng, J.; Keller, B. O.; Li, L. *J. Proteome Res.* **2003**, *2*, 543-552.

(39) Chottiner, E. G.; Shewach, D. S.; Datta, N. S.; Ashcraft, E.; Gribbin, D.; Ginsburg, D.; Fox, I. H.; Mitchell, B. S. *Proc. Natl. Acad. Sci. U S A* **1991**, *88*, 1531-1535.

(40) Nakatsuru, S.; Sudo, K.; Nakamura, Y. *Cytogenet. Cell Genet.* **1995**, *68*, 226-230.

(41) Simon, H. U.; Mills, G. B.; Kozlowski, M.; Hogg, D.; Branch, D.; Ishimi, Y.; Siminovitch, K. A. *Biochem. J.* **1994**, *297 (Pt 2)*, 389-397.

(42) Diaz-Jullien, C.; Perez-Estevez, A.; Covelo, G.; Freire, M. *Biochim. Biophys. Acta* **1996**, *1296*, 219-227.

(43) Hapke, S.; Kessler, H.; Luber, B.; Benge, A.; Hutzler, P.; Hofler, H.; Schmitt, M.; Reuning, U. *Biol. Chem.* **2003**, *384*, 1073-1083.

(44) Akiyama, N.; Matsuo, Y.; Sai, H.; Noda, M.; Kizaka-Kondoh, S. *Mol. Cell. Biol.* **2000**, *20*, 3266-3273.

(45) Wilkin, F.; Savonet, V.; Radulescu, A.; Petermans, J.; Dumont, J. E.; Maenhaut, C. *J. Biol. Chem.* **1996**, *271*, 28451-28457.

(46) Wharton, S. B.; Hibberd, S.; Eward, K. L.; Crimmins, D.; Jellinek, D. A.; Levy, D.; Stoeber, K.; Williams, G. H. *Br. J. Cancer.* **2004**, *91*, 262-269.

(47) Ruutu, M.; Peitsaro, P.; Johansson, B.; Syrjanen, S. *Int. J. Cancer* **2002**, *100*, 318-326.

(48) Hubeek, I.; Peters, G. J.; Broekhuizen, A. J.; Talianidis, I.; Schouten van Meeteren,

A. Y.; van Wering, E. R.; Gibson, B.; Creutzig, U.; Kaspers, G. J. *Nucleosides Nucleotides Nucleic Acids* **2004**, *23*, 1351-1356.

(49) Cao, Y.; Dave, K. B.; Doan, T. P.; Prescott, S. M. *Cancer Res.* **2001**, *61*, 8429-8434.

(50) Kinnula, V. L.; Lehtonen, S.; Sormunen, R.; Kaarteenaho-Wiik, R.; Kang, S. W.; Rhee, S. G.; Soini, Y. *J. Pathol.* **2002**, *196*, 316-323.

(51) Yokota, S.; Yanagi, H.; Yura, T.; Kubota, H. *J. Biol. Chem.* **1999**, *274*, 37070-37078.

(52) Maul, R. S.; Song, Y.; Amann, K. J.; Gerbin, S. C.; Pollard, T. D.; Chang, D. D. *J. Cell Biol.* **2003**, *160*, 399-407.

(53) Razzaque, M. S.; Taguchi, T. *Histol. Histopathol* **1999**, *14*, 1199-1212.

(54) Hebert, C.; Norris, K.; Della Coletta, R.; Reynolds, M.; Ordonez, J.; Sauk, J. J. *J. Cell. Biochem.* **1999**, *73*, 248-258.

(55) Helwani, F. M.; Kovacs, E. M.; Paterson, A. D.; Verma, S.; Ali, R. G.; Fanning, A. S.; Weed, S. A.; Yap, A. S. *J. Cell Biol.* **2004**, *164*, 899-910.

(56) Shapland, C.; Hsuan, J. J.; Totty, N. F.; Lawson, D. *J. Cell Biol.* **1993**, *121*, 1065-1073.

(57) Shields, J. M.; Rogers-Graham, K.; Der, C. J. *J. Biol. Chem.* **2002**, *277*, 9790-9799.

(58) Li, Y.; Lin, J. L.; Reiter, R. S.; Daniels, K.; Soll, D. R.; Lin, J. J. *J. Cell Sci.* **2004**, *117*, 3593-3604.

(59) Johnson, S.; Michalak, M.; Opas, M.; Eggleton, P. *Trends Cell Biol.* **2001**, *11*, 122-129.

(60) Yamada, A.; Irie, K.; Hirota, T.; Ooshio, T.; Fukuhara, A.; Takai, Y. *J. Biol. Chem.* **2005**, *280*, 6016-6027.

(61) Lu, Z.; Ghosh, S.; Wang, Z.; Hunter, T. *Cancer Cell* **2003**, *4*, 499-515.

(62) Galbiati, F.; Volonte, D.; Brown, A. M.; Weinstein, D. E.; Ben-Ze'ev, A.; Pestell, R. G.; Lisanti, M. P. *J. Biol. Chem.* **2000**, *275*, 23368-23377.

(63) Reynolds, A. B.; Carnahan, R. H. *Semin. Cell Dev. Biol.* **2004**, *15*, 657-663.

(64) Kinnula, V. L.; Paakko, P.; Soini, Y. *FEBS Lett.* **2004**, *569*, 1-6.

(65) Hajra, K. M.; Fearon, E. R. *Genes Chromosomes Cancer* **2002**, *34*, 255-268.

(66) Kawanishi, J.; Kato, J.; Sasaki, K.; Fujii, S.; Watanabe, N.; Niitsu, Y. *Mol. Cell. Biol.* **1995**, *15*, 1175-1181.

(67) Hayashi, E.; Kuramitsu, Y.; Okada, F.; Fujimoto, M.; Zhang, X.; Kobayashi, M.; Iizuka, N.; Ueyama, Y.; Nakamura, K. *Proteomics* **2005**, *5*, 1024-1032.

(68) Misek, D. E.; Imafuku, Y.; Hanash, S. M. *Pharmacogenomics* **2004**, *5*, 1129-1137.

(69) Wheelock, M. J.; Johnson, K. R. *Annu. Rev. Cell Dev. Biol.* **2003**, *19*, 207-235.

(70) Jang, J. S.; Cho, H. Y.; Lee, Y. J.; Ha, W. S.; Kim, H. W. *Oncol. Res.* **2004**, *14*, 491-499.

Chapter 5

Differential Dimethyl Labeling of N-termini of Peptides after Guanidination for Proteome Analysis [a]

5.1 Introduction

Stable isotopic labeling of proteins or peptides, in combination with liquid chromatography (LC) and mass spectrometry (MS), has increasingly been used in the past several years for quantitative proteome analysis [1-7]. Among these methods, the isotope-coded affinity tag (ICAT) approach pioneered by Aebersold and coworkers [1, 8-10] has been extensively used. The main advantage of this method is that it enriches peptides containing the rare amino acid cysteine, thereby significantly reducing the complexity of the peptide mixture and increasing the dynamic range of MS analysis [11, 12]. On the other hand, the use of the ICAT reagents fails for quantification of cysteine-free proteins. In addition, the ICAT reagents are structurally complex and thus the cost of the reagents is high.

---

[a] A portion of this chapter will be published as: C. Ji, N. Guo and L. Li, "Differential Dimethyl Labeling of N-termini of Peptides after Guanidination for Proteome Analysis" *J.Proteome Res.* **In press**. Ms. N. Guo collected all of the MALDI MS spectra using Bruker MALDI TOF and part of the MALDI MSMS spectra using an Applied Biosystems/MDS-Sciex QSTAR Pulsar QqTOF instrument.

As alternatives to ICAT, other labeling protocols have been developed including metabolic labeling of proteins in growing cells [3, 6, 13-17] and chemical labeling of peptides after protein digestion [2, 5, 18-26]. The main advantage of a metabolic labeling method is that the stable isotopes are incorporated into proteins at an early stage of the proteome study. Thus any variations in downstream protein sample workup will not significantly affect the quantitative results. However, this approach is precluded in quantitative analysis of tissues and biological body fluids, and human subjects. Cost is another negative aspect of such labeling approaches.

Chemical labeling is a more universal approach. In addition to providing quantitative information about proteomes, another important aspect of the chemical labeling approach is the capability to encode chemical or sequence information about the peptides to be detected and identified. The encoded information can greatly facilitate the peptide identification and improve the confidence of protein identification. For example, database searching combined with peptide MS/MS data generated from mass spectrometry has been widely used for protein identification. The efficiency of this protein identification approach ultimately requires that the recorded MS/MS spectra be matched accurately to peptide sequences from the corresponding database entries inferred from known or predicted gene sequences using bioinformatics tools. However, in most cases, fragment ion patterns shown in MS/MS spectra tend to be incomplete. They can also be very complex due to many factors, including the presence of neutral loss ions, contaminants, or noise peaks [21, 27]. In addition, many novel proteins resulting from co- or post-translational modifications (PTMs) or gene products undetected by the genomic or bioinformatic tools (e.g. via gene splicing or mutations) are not encoded in DNA sequences [28]. Furthermore, the large number of proteins present in a proteome also give a high degree of mass degeneracy in the protein database, resulting in false-positive matches in error-tolerant genome-dependent searches [27]. Consequently, any techniques that can encode chemical or sequence information to the fragment ions will significantly enable the protein identification processes.

To this end, various isotope labeling methods have been investigated for the production of mass-tagged peptides that can be used to facilitate database searching or spectral interpretation. For example, $H_2{}^{18}O$ can be used in protease digestion to introduce the ^{18}O tag through the hydrolysis reaction to label all proteolytic peptides uniformly at the C-terminus [29-31], allowing the y-type ions to be distinguished from the b-ions. Similarly, C-terminal esterification [32, 33] or N-terminal sulfonation [34] can also be used to tag peptides. Modification by nicotinyl-n-hydroxysuccinimide at the N-terminus has been used to label proteins [20]. Lysine-containing peptides can be characterized by using mass-coded abundance tagging (MCAT) [21]. In vivo mass tagging with various isotope-labeled amino acids has also been used to facilitate the interpretation of MS/MS spectra [13, 14, 27, 35].

Recently, differential dimethyl labeling of N-termini and ε-amino groups of lysine residues of tryptic peptides with d(0)- or d(2)-formaldehyde combined with either LC-ESI or microbore LC-MALDI have been developed and used for qualitative and quantitative proteome analysis [19, 36-39]. For example, dimethyl labeling combined with LC-MALDI MS and MS/MS has been applied to determine the proteins differentially expressed between an E-cadherin-deficient human carcinoma cell line (SCC9) and its transfectants expressing E-cadherin (SCC9-E) [38]. A total of 5480 peptide pairs were examined and 320 of them showed relative intensity changes of greater than 2-fold which led to the identification of 49 differentially expressed proteins. More recently, dimethyl labeling was shown to be useful for *de novo* sequencing of neuropeptides from a crude extract of *Cancer borealis* pericardial organ, and sequences of 117 peptides including a number of novel peptides were unambiguously determined by LC-ESI MS/MS [39].

However, the reported differential dimethyl labeling method has the following shortcomings for bottom-up proteomics. First, the difference of 4 Da between peptide pairs with a mass of greater than 1900 Da leads to a significant overlap of the isotope envelopes for peptide pairs. Therefore, to generate accurate quantification results, additional peak deconvolution using an appropriate software algorithm is required, which complicates the data analysis process. Second, missed cleavages are observed very often when trypsin is used as the enzyme, which results in 1 to 3 lysine residues occurring frequently in the tryptic peptides. Thus, multiple labeling of tryptic peptides is often observed and it complicates the selection of peptide pairs for quantification and collision-induced dissociation (CID). This is particularly true for an LC-MALDI based method in which selection of proper peptide pairs is critical for abundance-ratio-dependent quantitative analysis.

In this study, a modified N-terminal dimethyl labeling strategy, in which the N-termini of tryptic peptides are differentially labeled with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde after lysine residues in peptides are blocked by guanidination, is presented. Guanidination is known to be effective for selective labeling of lysine residues in peptides [5, 18, 20, 21, 34, 40-42]. It is demonstrated that N-terminal dimethylation (2ME) after lysine guanidination (GA) or 2MEGA provides uniform 6-Da differential isotope tags on peptides, which facilitates protein identification and quantification. To illustrate the applicability of this approach for proteome analysis, it is shown that 2MEGA labeling and microbore LC-MALDI MS can be used to identify polymorphic variants and low abundance proteins in complex samples, such as the whey fraction of bovine milk. In addition, by analyzing a mixture of differentially labeled protein digests from two equal amounts of whey fractions, it is demonstrated that this method can accurately determine the relative amounts of the proteins in these two control samples.

5.2 Experimental

5.2.1 Chemicals and Reagents

d(0),^{12}C-Formaldehyde (37% (w/w) in H_2O), O-methylisourea, sodium hydroxide, sodium bicarbonate, sodium cyanoborohydride, horse myoglobin, bovine trypsin and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (Oakville, ON, Canada). d(2),^{13}C-formaldehyde (~20% (w/w) in deuterated water) was obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). Acetonitrile was purchased from Fisher Scientific Canada (Edmonton, AB, Canada). Water used in these experiments was obtained from a Milli-Q Plus purification system (Millipore, Bedford, MA). The MALDI matrix, 2,5-dihydroxybenzoic acid (DHB), was purchased from Aldrich (Milwaukee, WI). α-Cyano-4-hydroxycinnamic acid (HCCA) and sodium dodecylsulfate (SDS) were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). HCCA was recrystallized from ethanol (95%) at 50 °C prior to use.

5.2.2 Tryptic Digestion of Myoglobin

A 100 µL of 0.5 µg/µL myoglobin solution was denatured by incubating at 95 °C for 10 min. After cooling down to room temperature, pH of the solution was adjusted to ~8.5 by adding 10 µL 1 M $NaHCO_3$. Then myoglobin was digested with trypsin overnight at 37 °C at an enzyme/substrate ratio of 1:50 (w/w). The digests were stored at –78 °C until use.

5.2.3 Preparation and Digestion of Whey Fractions of Skim Milk

Skim milk (30 µL containing ~1 mg proteins), purchased from a local grocery store, was diluted 1:50 (v/v) with Milli-Q water. Ultrafiltration of two equal volumes of diluted skim milk using Mirocon-3 (Millipore) with 3000 Da molecular mass cut-off was performed to remove carbohydrates, vitamins and salts. Whey protein fractions were prepared as reported in the literature with some modification [43]. Caseins (about 4/5 of total proteins) were precipitated by acidification of skim milk to pH 4.6 with 2 M acetic acid and removed by centrifugation (15,000 × g, 10 min). The supernatants were transferred into two new 2 mL siliconized vials and dried down to ~100 µL. Then the disulfide bonds of proteins in each vial were reduced with dithiothreitol (DTT). The free thiol groups were blocked by reaction with iodoacetamide, and the reduced and alkylated proteins were acetone precipitated to remove unreacted DTT and iodoacetamide. Acetone was pre-cooled at –78 °C and added to the protein solutions to a final concentration of 83% (v/v). After vortexing for 1 min, the mixture was kept at –20 °C for 2 h and then centrifuged for 10 min at 15,000 × g at 4 °C. The supernatant was decanted and properly disposed. Acetone was evaporated by incubating the pellet at room temperature. For the in-solution digestion, each pellet was solubilized by the addition of 5 µL 1% SDS (w/v) and the mixture was vortexed thoroughly. The SDS concentration was diluted to 0.05% and the pH was adjusted to ~8.5 with 10 µL 1 M $NaHCO_3$. Finally, 2.5 µL of 1 µg/µL bovine trypsin and 5 µL of 20 mM $CaCl_2$ were added to the mixture and allowed to digest overnight at 37 °C. The digests were kept at –78 °C until use.

5.2.4 2MEGA Labeling

Guanidination of lysine residues was performed as described in previous literature [5, 18, 40, 41] with some modification. Trypsin in both myoglobin digest and whey fraction digest solutions were irreversibly inactivated by adding 10 μL 2 M sodium hydroxide. The ε-amino groups of all lysines were blocked by adding 40 μL 2 M O-methylisourea in 100 mM $NaHCO_3$, adjusting to pH 11 with 2 M sodium hydroxide and incubating the resultant mixture at 37 °C for 2 h or 65 °C for 10 min. Then the reaction was stopped and the pH adjusted to 8 by adding 10% TFA. Reductive methylation with either d(0), ^{12}C-formaldehyde or d(2), ^{13}C-formaldehyde was also carried out as described previously [19, 36-38] with some modification. Each of the above guanidinated peptide solution was mixed with 15 μL 2 M sodium cyanoborohydride. The mixture was then vortexed and mixed with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde (4% (w/w) in water, 3 μL). The mixtures were vortexed and incubated at 37 °C for 1h. If necessary, ammonium bicarbonate (1 M, 3 μL) was added to consume the excess formaldehyde. **Cautions**: sodium cyanoborohydride is a highly toxic compound that releases hydrogen cyanide gas upon exposure to strong acid and formaldehyde is known to have carcinogenic effects including the unit cancer risk for inhalation exposure. Therefore, N-terminal dimethyl labeling process must be performed in a fume hood.

5.2.5 MALDI MS

The MALDI MS data of native, guanidinated, d(0),^{12}C- and d(2),^{13}C-formaldehyde labeled tryptic digests of horse myoglobin were obtained using a Bruker Reflex III MALDI time-of-flight (TOF) mass spectrometer (Bremen/Leipzig, Germany) equipped with a SCOUT 384 multiprobe inlet and a 337 nm nitrogen laser operated with a 3 ns pulse in positive ion mode with delayed extraction using reflectron mode. The two-layer sample deposition method [44] with HCCA as matrix was used in the MALDI MS analysis. The first layer was prepared as a 20 mg/mL HCCA solution in 20% methanol/acetone HCCA and the second layer with a saturating solution of matrix in 30% (v/v) methanol/water. The second layer was mixed with the diluted and acidified peptide solution at a ratio of matrix to analyte of 4:1 and the mixture was vortexed. After 0.5 μL of the first layer was deposited on the sample probe and air-dried, 0.5 μL of a mixture of the sample and the second layer solution containing 5 ng myoglobin digest was deposited on top of the first layer, allowed to air-dry and washed three times with 1 μL water each time. The sample spot was scanned with the laser beam under video observation and spectra were acquired by averaging 300-500 individual laser shots and processed with the Bruker supporting software. The spectra were externally calibrated with MALDI-MS standard peaks. The data were then reprocessed using the Igor Pro software package (WaveMetrics, Lake Oswego, Oregon, USA).

5.2.6 Microbore LC-MALDI QqTOF Mass Spectrometric Analysis

The separated d(0),^{12}C and d(2),^{13}C-formaldehyde labeled peptide mixtures from the whey fractions of the skim milk were combined and the resultant peptide mixture was desalted using a 2.1 × 12.5 mm RP-LC guard column (Zorbax, SB-C8, 5μm, Part No.: 821125915, Agilent, Palo Alto, CA). The eluate was dried down to ~100 μL with a SpeedVac. Then the desalted peptide mixture was separated by reversed-phase (RP) chromatography on a 1.0 × 150 mm Vydac C18 column (Catalog No.: 218TP5115, 5μm) at a flow rate of 40 μL/min with an Agilent 1100 series capillary HPLC equipped with an autosampler (Palo Alto, CA). A 2.1 × 10 mm Javelin style SDS removal cartridge column (Catalog No.: J2SDS, Western Analytical Products, Inc., Murrieta, CA) was added before the Vydac C18 column. Gradient elution was performed with solvent A (Milli-Q water, 0.1% TFA and 4% acetonitrile, v/v/v) and B (0.1% TFA in acetonitrile, v/v). The gradient was 0-10% B in 5 min, 10-40% B in 80 min, 40-90% B in 15 min. About 60 μg N-terminal dimethyl labeled digests were injected. 1 min HPLC fractions were directly collected onto a 100-well MALDI plate (Applied Biosystems, Concord, ON, Canada) using a home-built, heated droplet LC-MALDI interface [45]. After the fractionation was completed, the dried peptides in each well were redissolved and mixed with DHB matrix by the addition of 0.8 μL of 100 mg/mL DHB matrix in 50%ACN/50%water (v/v).

MALDI MS and MALDI MS/MS data were acquired on an Applied Biosystems/MDS-Sciex QSTAR Pulsar QqTOF instrument equipped with an orthogonal MALDI source employing a 337 nm nitrogen laser (Concord, ON, Canada) that has been previously described [37]. The instrument was operated in positive ion mode and collision-induced dissociation (CID) of peptides was achieved with argon as collision gas. Spectra were acquired and processed using Sciex supporting software and re-processed with Igor Pro software (WaveMetrics, Lake Oswego, Oregon, USA) for presentation.

5.2.7 Protein Identification from MS/MS Data

Peptide sequences were identified by database searching of the MS/MS spectra using the Mascot search program (http://www.matrixscience.com). Database searching was restricted to Mammalia in Swiss-Prot database. S-acetamido, N-terminal and lysine modifications were selected as fixed, methionine oxidation and threonine/serine phosphorylation as variable, 4 missed cleavages allowed and mass tolerance of the precursor peptide and its fragments set at ± 0.3 Da. In all cases, MS/MS data were searched twice; in one case with no requirement that the peptides be tryptic, the other being constrained to only tryptic peptides. The search results were confirmed by comparing the fragment ion peak matches using information derived from the isotope tag at the N-termini of the peptides. For those good quality spectra that failed to produce matches, de novo sequencing was done by comparing the spectra of d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde labeled peptide pairs (see Results and Discussion). The BLAST program of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) was used for sequence validation.

5.2.8 LC-ESI MS

Isotope effect analysis was carried out in a Bruker/Agilent Esquire-LC Ion Trap LC/MS^n system as previously described [37]. Myoglobin was initially digested to peptides with trypsin. Then the ε-amino groups on the side chain of lysine residues were blocked by guanidination and the solution was divided into two equal aliquots, each of which was dimethylated separately using either d(0),^{12}C- or d(2),^{13}C-formaldehyde. After labeling, the d(0),^{12}C- or d(2),^{13}C-formaldehyde labeled myoglobin digests were combined and the mixture was separated by RP-HPLC. The HPLC pump was operated at a flow rate of 100 μL/min and split to obtain flow through a microbore 150 × 1 mm i.d. Vydac C18 column at 40 μL/min. As peptides eluted from the column, they were electrosprayed directly into the ion trap mass spectrometer. Solvent delivery and separations were performed on an Agilent HP1100 HPLC system (Palo Alto, CA). Gradient elution was performed with solvent A (Milli-Q water, 0.1% TFA and 4% acetonitrile, v/v/v) and B (0.1% TFA in acetonitrile, v/v). The gradient was 0-10% B in 10 min, 10-40% B in 40 min, 40-90% B in 15 min. Mass spectra were acquired over the mass range m/z 400–1800. Reconstructed ion chromatograms of differentially labeled peptide pairs were obtained using Bruker Daltonics DataAnalysis software. The time shift between N-terminal differential dimethyl labeled pairs was calculated by comparison of elution time at the peaks of extracted ion chromatograms of the corresponding pair.

5.3 Results and Discussion

5.3.1 2MEGA Labeling Strategy

Both guanidination and reductive amination are well-known organic reactions that are used extensively in the modification of peptides [5, 19, 21, 22, 37] and proteins [40]. In this study, the N-terminal differential dimethyl labeling method relies on two steps (see Figure 5.1): first, the selective and quantitative (i.e., complete) blocking of ε-amino groups of lysine residues of tryptic peptides via guanidination at high pH with *O*-methylisourea (guanidination); second, after guanidination, specifically and quantitatively differential dimethyl labeling of N-termini of tryptic peptides was achieved with either d(0),^{12}C- or d(2),^{13}C-formaldehyde, adding 28 or 34 Da to each peptide respectively (see Figure 5.1A). Guanidination efficiently transforms lysine into homoarginine, which is 42 Da heavier than lysine, but does not affect the peptide amino terminus or other side groups [21]. This selectivity is likely raised from the steric effect of the side chain from the N-terminal amino acid which significantly reduces the

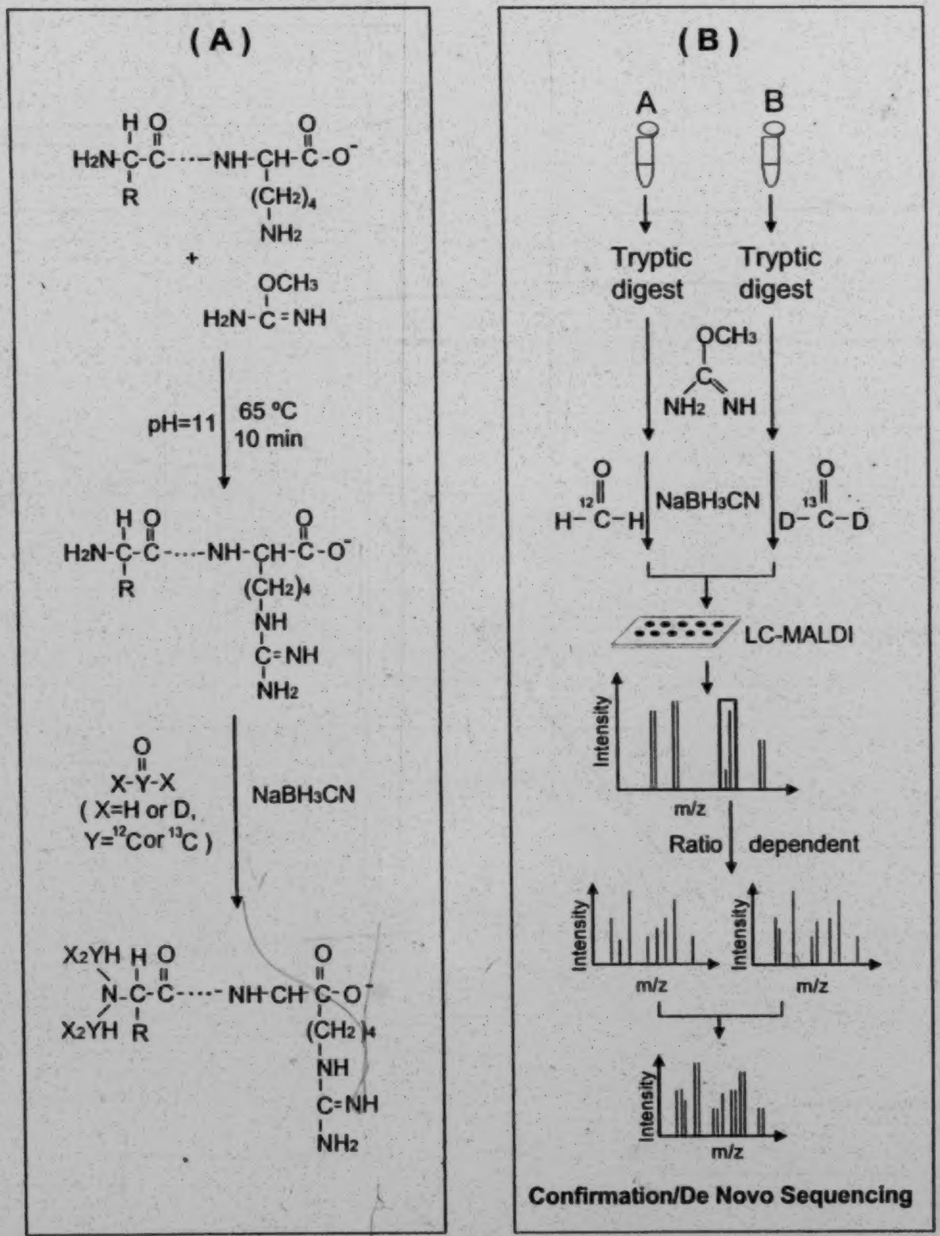
(A)
pH=11
65 °C
10 min
NaBH3CN
(X=H or D,
Y=12C or 13C)
(B)
A
B
Tryptic digest
Tryptic digest
NaBH3CN
LC-MALDI
Intensity
m/z
Ratio
dependent
Confirmation/De Novo Sequencing

Figure 5.1 Overview of the 2MEGA differential isotope labeling strategy for quantitative proteomic analysis and peptide sequencing. (A) The guanidination reaction selectively modifies the ε-amino group of the lysine residue and the N-terminus of the resultant peptide is isotopically labeled with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde. (B) Workflow for quantitative protein analysis and de novo peptide sequencing by combining the 2MEGA differential isotope labeling with microbore LC-MALDI QqTOF MS.

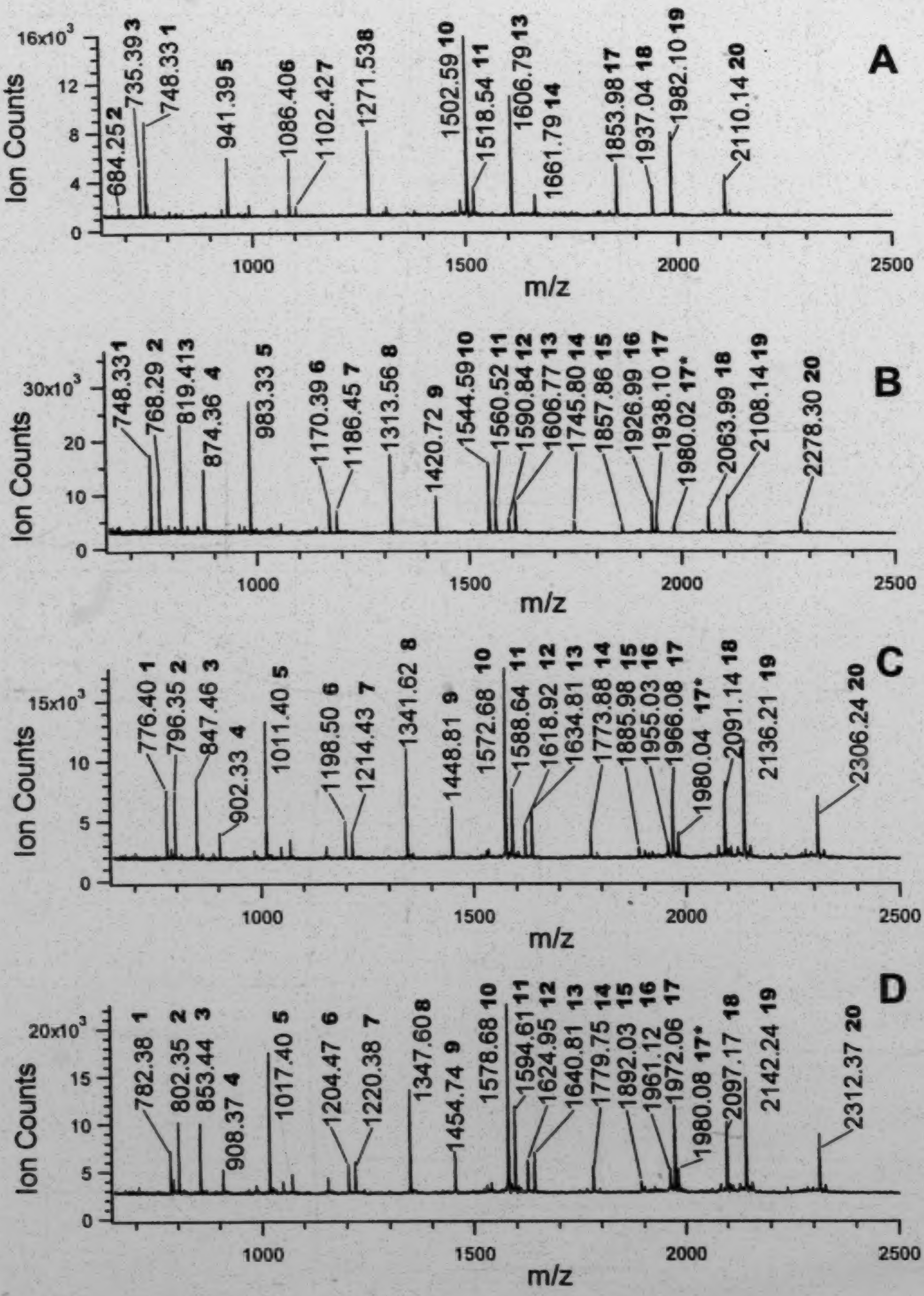

Figure 5.2 MALDI spectra of (A) native, (B) guanidinated, (C) N-terminal d(0),^{12}C-formaldehyde labeled, and (D) N-terminal d(2),^{13}C-formaldehyde labeled myoglobin digest.

Table 5.1 Theoretical and experimental MALDI-MS data for myoglobin tryptic digest*

Peak #	Sequence range	No. of lysines	Unmodified Mass**		Modified Mass					
					Lysine guanidinated		N-terminal dimethylated $(+H_4{}^{12}C_2)$		N-terminal dimethylated $(+D_4{}^{13}C_2)$	
			Theo.	Expt. (Fig.5 2A)	Theo.	Expt. (Fig.5.2B)	Theo.	Expt. (Fig.5.2C)	Theo.	Expt. (Fig.5.2D)
1	134-139	0	748.43	748.33	748.44	748.33	776.47	776.40	782.50	782.38
2	43-47	2	684.37	684.25	768.41	768.37	796.44	796.35	802.48	802.35
3	97-102	2	735.48	735.39	819.53	819.41	847.56	847.46	853.60	853.44
4	57-63	2	790.43	—***	874.47	874.36	902.50	902.33	908.54	908.37
5	146-153	1	941.47	941.39	983.49	983.33	1011.52	1011.40	1017.55	1017.40
6	48-56	2	1086.56	1086.4	1170.60	1170.39	1198.63	1198.50	1204.67	1204.47
7	48-56	3	1102.55	1102.42	1186.60	1186.45	1214.63	1214.43	1220.67	1220.38
8	32-42	1	1271.66	1271.53	1313.68	1313.56	1341.71	1341.62	1347.74	1347.60
9	64-77	1	1378.84	—	1420.86	1420.72	1448.89	1448.81	1454.92	1454.74
10	119-133	1	1502.66	1502.59	1544.69	1544.59	1572.72	1572.68	1578.75	1578.68
11	119-133	1	1518.66	1518.54	1560.69	1560.52	1588.72	1588.64	1594.75	1594.61
12	64-78	2	1506.94	—	1590.98	1590.84	1619.01	1618.92	1625.05	1624.95
13	17-31	0	1606.85	1606.79	1606.86	1606.77	1634.89	1634.81	1640.92	1640.81
14	32-45	2	1661.85	1661.79	1745.89	1745.80	1773.92	1773.88	1779.96	1779.75
15	1-16	0	1815.90	—	1857.92	1857.86	1885.95	1885.98	1891.98	1892.03
16	103-118	1	1885.02	—	1927.04	1926.99	1955.08	1955.03	1961.11	1961.12
17	80-96	2	1853.96	1853.98	1938.00	1938.10	1966.03	1966.08	1972.07	1972.06
18	32-47	3	1937.01	1937.04	2063.09	2063.99	2091.12	2091.14	2097.15	2097.17
19	79-96	3	1982.05	1982.1	2108.13	2108.14	2136.16	2136.21	2142.19	2142.24
20	78-96	4	2110.15	2110.14	2278.24	2278.30	2306.27	2306.24	2312.30	2312.37
			(94/153) 61.5%		(147/153)96.1%		(147/153)96.1%		(147/153)96.1%	

* The data are a summary of the spectral results shown in Figure 5.2. All ion masses listed are based on their protonated forms.

** Theo.: theoretical mass; Expt.: experimental mass; Fig.: figure

*** Not found.

guanidination efficiency for the terminal amine, as compared to guanidination of the amine group in the lysine side chain. In addition to blocking the reactivity of the lysine side chain, eliminating the multiple labeling of peptides [18], guanidination has additional advantages related to the MS-based analysis of the peptides. Guanidination converts lysine residues to homoarginine, making them more basic, and therefore easily being detected by MS [46, 47].

To examine the labeling efficiency of guanidination of ε-amino groups of lysine residues and differential dimethyl labeling of N-termini of tryptic peptides after guanidination, a tryptic digest of myoglobin was chosen as a model peptide mixture. The results from the tryptic digest are shown in Figure 5.2 and Table 5.1. The MALDI MS spectrum in Figure 5.2A is from the digest without any labeling. Figure 5.2B shows the mass spectrum of guanidinated peptides. Figure 5.2C and 5.2D show MALDI MS spectra of guanidinated peptides followed by N-terminal dimethyl labeling with either d(0),^{12}C- or d(2),^{13}C-formaldehyde. The peptide sequences assigned to the peptide ions observed in the MALDI spectrum, based on a database search, are tabulated in Table 5.1. In summary, 15 peptide ions were observed in the MALDI spectrum of the unlabeled tryptic digest of myoglobin while 20 peptide ions were observed in each MALDI spectrum of the modified digest. Compared to the theoretical sequence of myoglobin, peptide sequence coverage found from these identified tryptic, guanidinated and then dimethylated peptides was up to 96.2%. Compared to the coverage from the unlabeled tryptic peptides, guanidination increased the peptide sequence coverage by 58%, with 5 additional tryptic peptides found from each of the labeled digests. The signal intensities of most of the guanidinated peptides are comparable to the unlabeled peptides except for some lysine-containing peptide ions that are only observed in the MALDI spectra after guanidination. This is because the conversion of lysines into homoarginines increases the basicity of the peptides. Consistent with previous reports [18, 41], partial guanidination of the N terminus of the peptide that have glycine as the N-terminal residue was also observed (peak 17* in Figure 5.2B-D). This is understandable considering that glycine does not have a bulky side chain as other amino acids. Thus the steric effect on guanidination is relatively small, which results in the reduction of selectivity and partial guanidination of the terminal amine. For the other peptides observed, none of which have an N-terminal glycine, there is no evidence of guanidination at their N termini. In this study, I also tried to minimize this side reaction by keeping the reaction temperature at 37 °C or lower. Since an incubation time of 2 h was required for the reaction to be complete at 37 °C, compared with 10 min required when a high temperature of 65 °C was used, with no observed reduction in side reactions, all other guanidination reactions were carried on at 65 °C for 10 min.

From Figures 5.2B-D, one can see that no unlabeled starting peptides or incompletely modified products are observed in the MALDI spectra, which indicates that the yields of both guanidination of lysine residues with *O*-methylisourea and dimethyl labeling of N-termini of tryptic peptides with either d(0),^{12}C- or d(2),^{13}C-formaldehyde are quantitative. It was also noticed that all peptide ions observed in the MALDI spectrum of guanidinated peptides (Figure 5.2B) were also seen in the MALDI spectra of peptides labeled with guanidination and N-terminal dimethylation. The reason for this is that dimethyl labeling of N-termini of tryptic peptides does not significantly change the basicity of the N-termini of the peptides. Another advantage of the 2MEGA dimethyl labeling method is that guanidination and dimethyl labeling are chemically and procedurally compatible. Although two step reactions are used, these reactions can be done in a single vial without any inconvenient steps between them, such as transferring from one tube to another one, desalting or drying, which may result in sample loss or reduce the accuracy of quantification. In summary, the 2MEGA labeling reaction is simple, complete (i.e., quantitative), and easy to carry out under mild reaction conditions. In addition, the signal quality of the 2MEGA labeled peptide ions is enhanced compared with that of the unlabeled peptide ions.

5.3.2 Isotope Effect

The use of deuterium as the stable isotope label results in a primary isotope effect that can cause differential elution between the deuterated and non-deuterated peptides during RP-LC and may compromise measurement precision and accuracy [48, 49]. In this study, I have examined the isotope effect resulting from differential dimethyl labeling of the N-termini of peptides with d(0),^{12}C- or d(2),^{13}C-formaldehyde. I found the average peak separation of d(0),^{12}C- and d(2),^{13}C-formaldehyde labeled peptides was ~1 s on peaks that were 20-25 s wide at half-height (data not shown), which is negligible.

This can be explained by the following two reasons. First, the deuterium tags are added to the charged sites of peptides, which do not interact strongly with the non-polar stationary phase [50]. Thus the addition of the tags does not cause signification changes in retention behaviors of d(0) and d(2) labeled peptide pairs. Second, although a mass shift of 6 Da between the N-terminal dimethyl labeled peptide pairs was generated, only 4 deuterium atoms were added to one form of the peptide pair which also helps to reduce the isotopic effect. It should be noted that although the isotope effect is negligible for the N-terminal differential labeling strategy, accurate quantification should still be calculated using the summed monoisotopic peak intensities of the peptide pairs from the entire elution profile (see below).

5.3.3 CID of the 2MEGA Labeled Peptides

The 2MEGA labeled tryptic digest of myoglobin was further investigated by MALDI MS/MS for its applicability for protein identification by sequence analysis. Figure 5.3 depicts the fragment ion spectra obtained from the myloglobin digests of unlabeled (Figure 5.3A), N-terminal d(0),^{12}C-formaldehyde labeled (Figure 5.3B) and N-terminal d(2),^{13}C-formaldehyde labeled (Figure 5.3C) peptides (YKELGFQG). I note two interesting features in Figure 5.3 by comparing the CID fragments of the native and N-terminal differential dimethyl labeled tryptic peptides. First, simple and easily assignable fragment ions are clearly shown in MALDI MS/MS spectra of the N-terminal differential dimethyl labeled peptide compared to that of the native peptide. Second, the

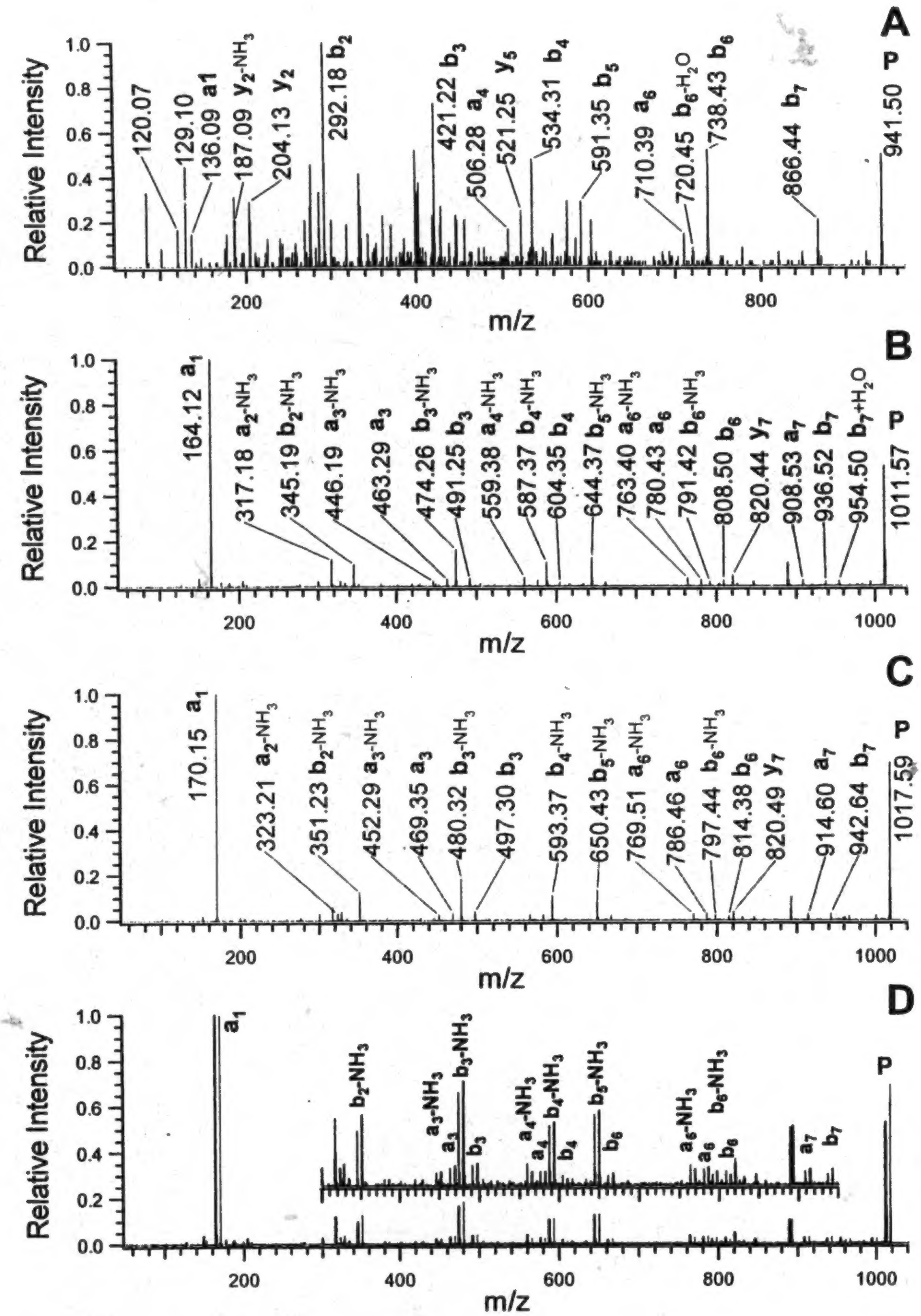

Figure 5.3 Selected MALDI MS/MS spectra of (A) native, (B) N-terminal d(0),^{12}C-formaldehyde labeled, (C) N-terminal d(2),^{13}C-formaldehyde labeled tryptic peptide YKELGFQG derived from myoglobin, and (D) Overlaid MS/MS spectra of (B) and (C).

signal intensities of the a_1 ions are increased substantially upon labeling (> 10 times), making it easy to be differentiated from the immonium ion peaks. Without labeling, these ions were hardly detectable or difficultly differentiable from the immonium ion peaks of other amino acids in the MALDI MS/MS spectra of the unlabeled digests of many proteins. This phenomenon may be due to the fact that the resulting b_1 ion tends to form a stable alkylated immonium ion a_1 by losing a molecule of carbon monoxide [36]. It is also noticeable that, for some fragment ions, the number of b ions detected was significantly increased upon labeling. The data obtained proves the applicability of N-terminal dimethyl labeling for the analysis of peptide sequences using the MALDI MS/MS technique and suggest that this method is likely to have a greater degree of success because of enhanced a_1 and easily assignable fragment ions relative to methods using unlabeled counterparts.

5.3.4 Quantification and Identification of Phosphorylated Peptides and Low Abundance Proteins in Milk

The feasibility of using the combination of the N-terminal dimethyl labeling and microbore LC-MALDI for quantification and identification of phosphorylated peptides and low abundance proteins in a mixture was further investigated using skim milk as the study model. Skim milk was chosen as it represents a biological fluid that metabolic labeling cannot be used for quantitative proteome analysis. In addition, skim milk is considerably less complex than serum, but it contains several high abundance proteins which presents a challenge in detecting the low abundance proteins. Furthermore, the proteome of bovine milk has not been well characterized [51] and, thus, the information generated from this work should expand my knowledge on the milk proteome.

Figure 5.1B shows the workflow for quantitative comparison of proteins in two samples in a single LC-MALDI analysis. Two whey fractions were prepared from two equal volumes of skim milk samples following the procedure described in the Experimental Section. After tryptic digestion, peptides from each fraction were guanidinated, then N-terminal dimethyl labeled with d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde, respectively. Finally, the N-terminal differential dimethyl labeled peptides from the two whey fractions were combined and examined by microbore LC-MALDI MS to quantify proteins by comparing the relative signal intensities of monoisotopic peaks of peptide pairs with a mass shift of 6.03 Da, and MS/MS analysis followed by database searching or manual interpretation to identify phosphopeptides and proteins. As previously reported [9, 18, 37], LC-MALDI based quantitative methods allow abundance ratio dependent analysis. Here, since the theoretical ratios for all the observed peptides should be equal to 1, all the observed peptide pairs with reasonably good signal-to-noise ratios (i.e., over ~4) were selected for CID. In summary, 56 peptide pairs were quantified and identified. From them, 14 proteins were unambiguously identified by using the sequence information derived from d(0),^{12}C- and d(2),^{13}C-formaldehyde labeled peptide pairs (see Table 5.2). Although this sample was digested with trypsin, for unknown reasons, the proteins in the mixture were also cleaved frequently at nonspecific cleavage sites. Nonspecific cleavage in analyzing other proteome samples has been reported [52]. It should also be noted that the guanidination reaction does not cause any problem to identify peptides with internal guanidinated lysine(s) using MALDI MS/MS. This is demonstrated in Table 5.2 where several peptides containing one or two internal lysines were identified.

The measured average ratio for all the peptide pairs is 1.02, which is very close to the theoretical ratio of 1.00. The calculated percentage error is 2.0% and relative standard deviation is 4.6%, well below 15%, normally treated as the threshold for quantitative bioanalysis. These results illustrate that the new 2MEGA differential isotope labeling method combined with LC-MALDI can provide accurate information on the relative abundances of proteins present in two samples. Real world applications to determine differentially expressed proteins in different cell lines (see Chapter 6), serum and tissue samples will be reported in the future.

It is worth noting that, using the above strategy, several phosphorylated peptide pairs were also accurately quantified and unambiguously identified in the tryptic digest of the whey fraction of the bovine milk (see Table 5.2). Figure 5.4 shows an example of a phosphorylated peptide pair from two whey fractions of the bovine milk that was quantified and identified using the integrated strategy of 2MEGA differential isotope labeling and microbore LC-MALDI QqTOF MS. Figures 5.4A and B show the partial MALDI spectra of two consecutive LC fractions collected between 33 and 35 min during the LC gradient run.. The phosphopeptide pair at the m/z values of 2131.80 and 2137.84

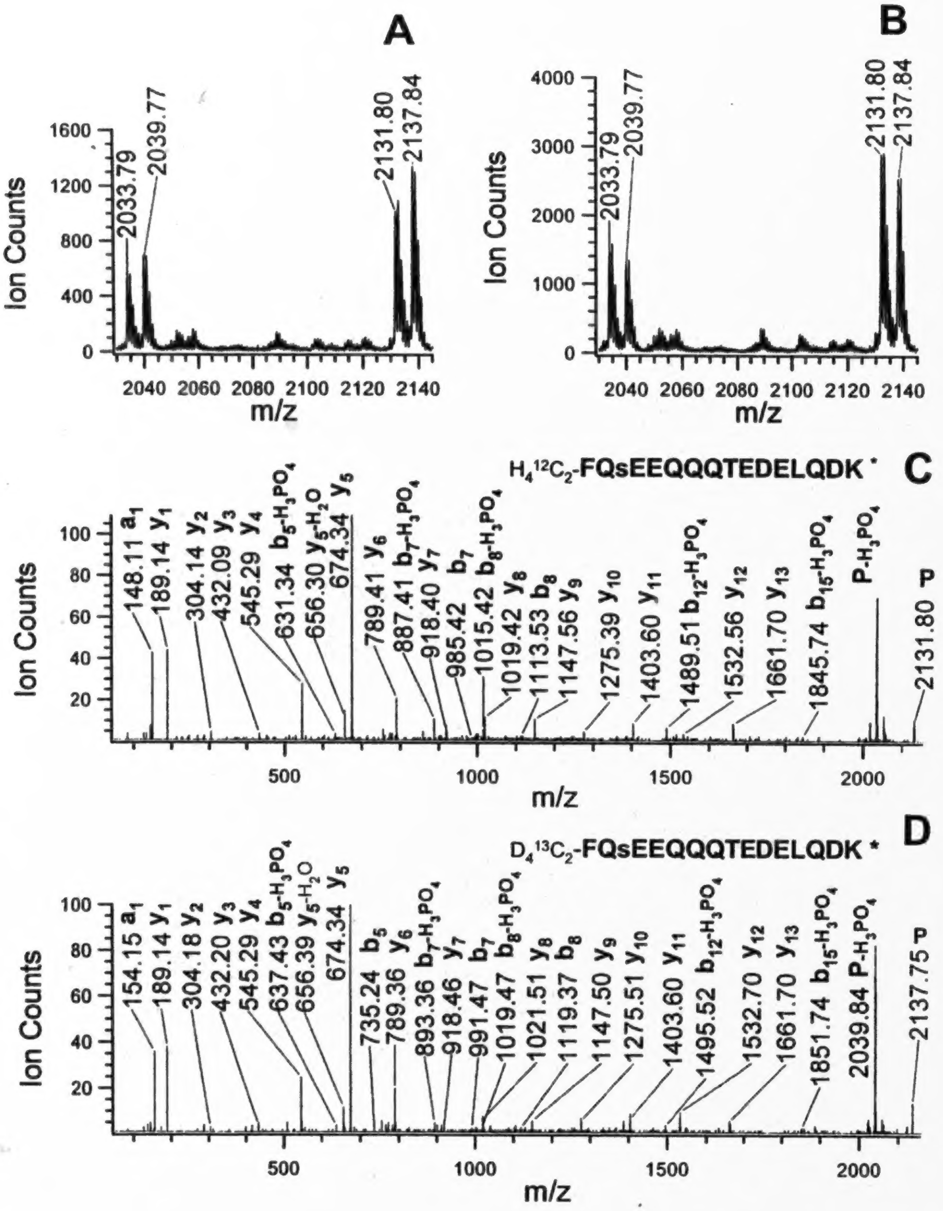

Figure 5.4 Accurate quantification and unambiguous identification of a phosphopeptide pair from a protein digest of bovine milk whey fraction. (A) and (B) Partial MALDI mass spectra obtained from two consecutive LC fractions on the MALDI target. (C) and (D) MALDI MS/MS spectra for a pair of peptides at m/z 2131.80 and 2137.84 (FQsEEQQQTEDELQDK).

Table 5.2 Identification and quantification of proteins in the whey fraction of skim milk.

#	Protein Name	Accession ID	Peptide Sequence*	Score	MW (kDa)
1	Beta-lactoglobulin	P02754	VLVLDTDYK	38	19.9
			IIAEKTK	22	
			LIVTQTMK	45	
			LIVTQTVK	de novo**	
			LIVTQTmK	35	
			LSFNPTQLEEQCHI	79	
			TPEVDDEALEK	78	
			SFNPTQLEEQCHI	38	
			VYVEELKPTPEGDLEILLQK	75	
			AMAASDISLLDAQSAPLR	56	
			VYVEELKPTPEGDLEIL	43	
			IPAVFKIDALNENK	39	
			TPEVDDEALEKFDK	19	
			TPEVDDEALEKFDKALK	26	
			ALPMHIR	32	
			SLAMAASDISLLDAQSAPLR	104	
			SLAmAASDISLLDAQSAPLR	69	
			TKIPAVF	22	
			YSLAMAASDISLLDAQSAPLR	22	
2	Serum albumin	P02769	LVNELTEFAK	64	69.3
			YICDNQDTISSK	74	
			TVMENFVAFVDK	29	
			AEFVEVTK	39	
			MPCTEDYLSLILNR	71	
			LVVSTQTALA	27	
			LGEYGFQNALIVR	77	
			LVTDLTK	27	
3	Alpha-S1 casein	P02662	HQGLPQEVLNENLLR	79	24.5
			SDIPNPIGSENSEK	36	
			YKVPQLEIVPNsAEER	37	
			KVPQLEIVPNsAEER	25	
			TTMPLW	25	
			FVAPFPEVFGK	20	
			VPQLEIVPNSAEER	53	
			VPQLEIVPNsAEER	53	

4	Lactotransferrin	P24627	SFQLFGSPPGQR	54	78.1
			GEADALNLDGGYIYTAGK	64	
			ECHLAQVPSHAVVAR	33	
			VVWCAVGPEEQK	29	
			ESPQTHYYAVAVVKK	24	
5	Lactophorin	P80195	LPLSILKEK	33	17.2
			EQIVIR	26	
			LPLSILK	44	
6	Alpha-S2 casein	P02663	NAVPITPTLNR	18	26.0
			TVDmEsTEVFTK	70	
			TVDMEsTEVFTKK	44	
7	Kappa casein	P02668	SPAQILQWQVLSNTVPAK	59	21.3
8	Polymeric-immunoglobulin receptor	P81265	ALLDPSFFAK	53	82.4
9	Cyclic dodecapeptide	P22226	AVDQLNEQSSEPNIYR	26	17.6
10	Prostaglandin-H2 D-isomerase	O02853	SLGFTEEGIVFLPK	53	21.2
11	Beta-2-microglobulin	P01888	IQRPPK	de novo	13.7
12	Alpha-lactalbumin	P00711	LDQWLCEK	22	16.2
			VGINYWLAHK	45	
			LDQWLCEKL	27	
			FLDDDLTDDIMCVK	111	
13	Osteopontin-K	P31098	**IRIsHELDSASSEVN**	24	31.0
			IRIsHELDSASSEVN***	45	
14	Beta casein	P02666	GPFPILV	29	25.1
			VLPVPQK	37	
			LLYQEPVLGPVR	55	
			FQsEEQQQTEDELQDK	84	
			FQSEEQQQTEDELQDKIHPF	45	
			YPVEPFTER	de novo	

* Sequences in bold are the phosphorylated peptides. The lowercase letter "s" stands for phosphorylated serine, and lowercase "m" stands for oxidized methionine.

** The peptide sequences were generated from de novo interpretation of the overlaid MS/MS spectra.

*** This peptide was detected as doubly phosphorylated. Only one phosphorylation site was unambiguously assigned. The other was not assignable due to the poor quality of the MALDI MS/MS spectra.

was observed in those two consecutive fractions. The accurate abundance ratio of 1.01, which is very close to the theoretical ratio of 1.00, is achieved by calculating the abundance ratio of this peptide pair using the summed monoisotopic intensities of the differentially labeled peptide pair. However, the abundance ratios of the observed peptide pairs, if calculated using the monoisotopic peak intensities of the observed peptide pair in Figures 5.4A and B, are 0.76 and 1.14, respectively, both of which are a little far from the theoretical ratio of 1.00. This result indicates that accurate quantification should be calculated using the summed signal intensities of the peptide pair during the whole profile. One also can see that the phosphopeptide pair observed on each of the MALDI spectra (Figures 5.4A and B) is accompanied by another peptide pair with a mass difference of 98 Da, likely corresponding to the loss of a phosphoric acid group from the phosphopeptide pair. This feature was consistently observed for all the phosphopeptides identified in this study. Therefore, in the future, this feature can potentially be used to quantify and identify phosphopeptides selectively in the quantitative phosphoproteome analysis by using this integrated strategy. Figures 5.4C and D show that the phosphorylation site can be unambiguously assigned with the help of the 2MEGA differential isotope labeling. In this case, the peptide sequence is determined to be FQsEEQQQTEDELQDK with a phosphorylation site at the indicated serine residue.

The proteome of bovine milk is dominated by just six gene products (α-S1 casein, α-S2 casein, β-casein, κ-casein, β-lactoglobulin and α-lactalbumin) that constitute approximately 95% of bovine milk protein [51]. This makes the quantification and identification of low abundance proteins in milk very challenging. To date, only a few proteomic studies have succeeded in identifying minor protein components of bovine milk [53, 54]. Separation of caseins from the whey fraction in the bovine milk by acid precipitation can enhance the resolution of the 2D gels. 2D PAGE of the whey fraction has been used to analyze several isoforms of the low abundance protein lipocalin-type prostaglandin D synthase [53]. However, acid precipitation can not completely remove all caseins from milk, since casein proteins were still shown in 2D gels [53] and, in this work, some of these proteins after acid precipitation (see Table 5.2) were also identified. In another study, Yuki and coworkers used immunoabsorption techniques and 2D PAGE followed by microsequencing and MS to help in detection and identification of several low abundance proteins, including bovine serum albumin, lactoferrin and serotransferrin of bovine mature skim milk [54].

In Chapter 3, it has been demonstrated that the high sample loading capacity of microbore LC-MALDI, compared to the conventional approach of using a capillary LC column, gives rise to accurate protein quantification due to the increased signal-to-noise ratio for a protein mixture with a concentration dynamic range of as high as 1×10^4 [37]. In this study, the combination of 2MEGA differential isotope labeling and microbore LC-MALDI has been used to identify low abundance proteins in bovine milk (see Table 5.2). Eight low abundance proteins in addition to the high abundance proteins were unambiguously identified. Among them, two proteins (osteopontin-K and cyclic dodecapeptide) were identified for the first time in mature bovine milk. While the total number of proteins identified is low, to my knowledge, this still represents the first report illustrating the detection of these many low abundance proteins in milk. Further fractionation of the milk sample at the protein and peptide levels will undoubtedly increase the proteome coverage, but this is not the main focus of the present study.

5.3.5 *De Novo* Peptide Sequencing with 2MEGA Differential Isotope Labeling

Figure 5.1B also illustrates the process of *de novo* peptide sequencing using the 2MEGA differential isotope labeling. Briefly the steps involved are as follows: (i) the protein sample is initially digested with trypsin; (ii) lysine residues of tryptic peptides are blocked by guanidination with *O*-methylisourea; (iii) the guanidinated digest is split into two equal parts in two new vials; (iv) the peptides in each vial are then labeled with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde; (v) the peptide pools are combined and the resultant peptide mixture is analyzed by microbore LC-MALDI MS and the peptide pairs are selected for CID. (vi) the spectra of the peptide pairs are overlaid and compared to delineate the fragment ions derived from the N-terminal of the peptide (a and b ions), allowing amino acid sequence to be assigned based on a reference table of amino acid residue masses.

Theoretical masses of the a_1 ions derived from the twenty amino acid residues after N-terminal differential dimethyl labeling with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde are summarized in Table 5.3. Without labeling, a_1 ion signals in low mass region rarely show up or are indistinguishable from immonium ion signals from other amino acid residues in the peptide sequence, even if they show up in the collected MALDI MS/MS spectra (e.g., see Figure 5.3A). Therefore, no or limited useful sequence information can be read. For the labeled peptides, however, the enhanced signals for the a_1 ions in the low mass region, combined with accurate mass determination and the information obtained from the overlaid spectra of the peptide pair, can be used to fingerprint easily and unambiguously the N-terminal residue. Except for the *L (leucine) and *I (isoleucine) pair whose masses are indistinguishable under low energy CID conditions, the remaining eighteen a_1 ions can be easily distinguished using a modern mass spectrometer.

Figure 5.5 shows an example of using the strategy mentioned above for *de novo* sequencing of a peptide. No significant match was obtained (see the inset in Figures 5.5B and C) when a database search was carried on the MALDI MS/MS spectra of this peptide pair with m/z of 1165.58 and 1171.62 (Figures 5.5B and C). The full sequence assignment for this unknown peptide in the tryptic digest of the whey fraction of bovine milk can be deduced readily according to the information extracted from the overlaid spectra (Figure 5.5D). By referring to the Table 5.3, N-terminal tyrosine residue (Y) can be assigned right away, based on the enhanced a_1 ions (164.10 and 170.11). Other amino acid residues are also unambiguously assigned as PVEPFTER by the analysis of a series of b ions. Finally, the deduced peptide sequence (YPVEPFTER) was validated using the public BLAST program. This deduced peptide was identified to be a partial

Table 5.3 Theoretical masses of the a_1 ions derived from the twenty amino acid residues without and with N-terminal dimethyl labeling with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde.

N-terminal amino acid residues	Unmodified a_1	Modified a_1 mass	
Proteinogenic amino acid residues	mass	$+H_4{}^{12}C_2$	$+D_4{}^{13}C_2$
Alanine (A)	44.050	72.081	78.113
Arginine (R)	129.114	157.145	163.177
Asparagine (N)	87.006	115.037	121.069
Aspartic acid (D)	88.040	116.071	122.103
Cysteine (C)*	133.056	161.087	167.119
Glutamic acid (E)	102.056	130.087	136.119
Glutamine (Q)	101.071	129.103	135.135
Glycine (G)	30.034	58.066	64.098
Histidine (H)	110.072	138.103	144.135
Isoleucine (I)	86.097	114.128	120.160
Leucine (L)	86.097	114.128	120.160
Lysine (K)	101.108	171.161	177.193
Methionine (M)	104.053	132.085	138.117
Phenylalanine (F)	120.081	148.113	154.144
Proline (P)	70.066	84.081	87.097
Serine (S)	60.045	88.076	94.108
Threonine (T)	74.061	102.092	108.124
Tryptophan (W)	159.092	187.124	193.155
Tyrosine (Y)	136.076	164.108	170.139
Valine (V)	72.081	100.113	106.144

* Side chain of cysteine was blocked by iodoactemide.

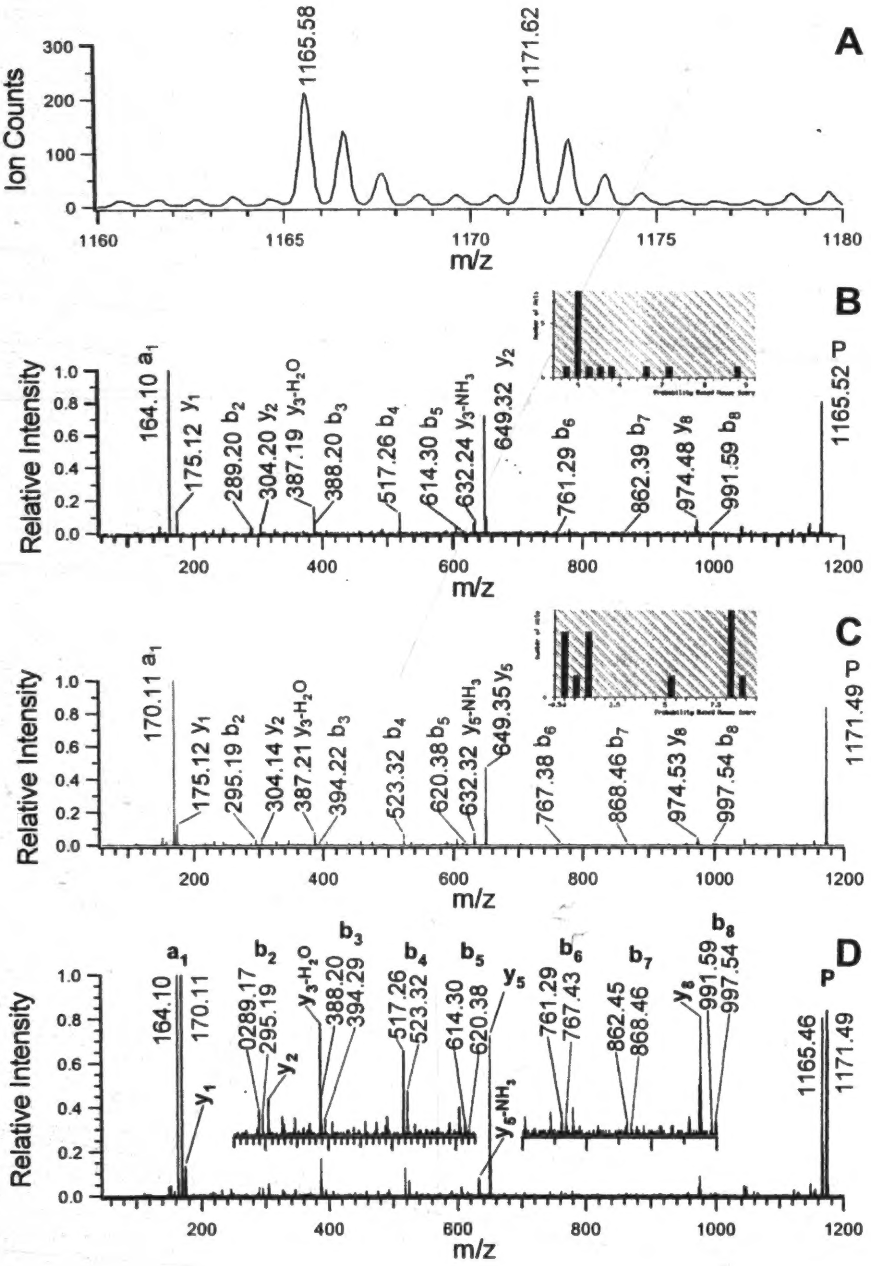

Figure 5.5 (A) Partial MALDI mass spectrum obtained from the fraction at 43 min from the microbore column separation of a protein digest of bovine milk whey fraction. (B) and (C) MALDI MS/MS spectra for a pair of peptides at m/z 1165.58 and 1171.62 along with the summary of MASCOT search results shown in insets. (D) Overlaid MS/MS spectra of (B) and (C).

sequence (129-137) of bovine β-casein. Note that, although β-casein and other caseins were precipitated out, the whey fraction apparently still contained a small amount of those proteins and the LC-MALDI technique with microbore column separation provided sufficient sensitivity to detect the peptides from those proteins. The whole sequence of the bovine β-casein in the Swiss-Prot database is as follows: MKVLILACLV ALALARELEE LNVPGEIVES LSSSEESITR INKKIEKFQS EEQQQTEDEL QDKIHPFAQT QSLVYPFPGP IPNSLPQNIP PLTQTPVVVP PFLQPEVMGV SKVKEAMAPK HKEMPFPK**YP** **VEPFTE**SQSL TLTDVENLHL PLPLLQSWMH QPHQPLPPTV MFPPQSVLSL SQSKVLPVPQ KAVPYPQRDM PIQAFLLYQE PVLGPVRGPF PIIV. The partial sequence of this protein matching with the deduced peptide sequence is in bold. On comparison, one can see that the serine residue at 137 is changed to arginine, making the peptide easily generated by trypsin and detected by MALDI MS. The detection of this single amino acid polymorphism in bovine β-casein is also consistent with that mentioned in the Swiss-Prot database. Thus, the failure of database searching of the labeled peptide pair can be explained and it might be due to the amino acid variation (S → R) that has occurred in the peptide sequence.

Although the detection of single amino acid polymorphism is very difficult because it requires the detection of the peptide in the MS scan and also complete sequence information in the tandem MS analysis, another new variant in β-lactoglobulin. Amino acid variation occurs at residue 23 (M → V) was successfully identified (see

Figure 5.6). Without the labeling experiment, determination of these single amino acid mutations would not be possible.

5.4 Conclusions

A chemical labeling strategy for quantitative and qualitative proteome analysis, based on N-terminal dimethylation after lysine guanidination to tag the N-termini of peptides selectively, have been developed. This 2MEGA labeling strategy exhibits the following important features. First of all, it incorporates a single mass tag in every peptide which is not N-terminally blocked, derived from any protein from any source: cell culture, tissue, or biological fluids. Second, the uniform 6 Da mass difference between light and heavy isotopically labeled peptide pairs eliminates overlap of the isotope envelopes even for peptide pairs with a mass of 3000 Da, the upper threshold to choose peptides for MS/MS analysis using most of the currently available instruments, and greatly simplifies the quantitative data analysis. Third, the 2MEGA differential isotope labeling is relatively inexpensive and can be done with commercially available reagents. It has high reaction efficiency under mild reaction conditions, and negligible isotope effect on reversed-phase separation. Finally, the presence of universal a_1 ions in the MALDI MS/MS spectra and the overlaid fragment ion spectra generated from a pair of differentially labeled peptides can be used to confirm peptide sequences obtained from MS/MS database searching or to carry out *de novo* sequencing of peptides based on their

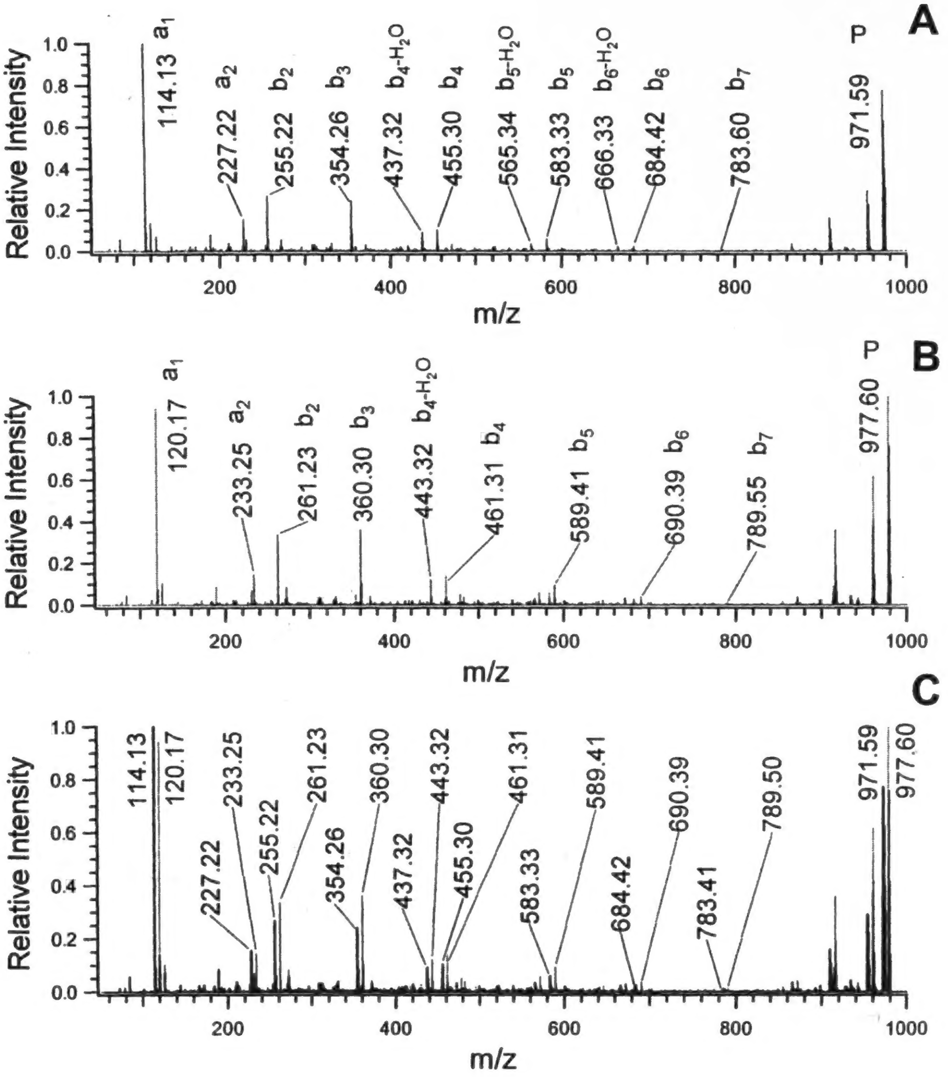

Figure 5.6 An example shows the identification of the new variant in β-lactoglobulin. The peptide sequence was de novo as L(I)I(L)VTQTVK, which corresponds to (17)LIVTQTMK(24) in β-lactoglobulin with amino acid variation occurring at residue 23 (M → V). (A) and (B) MALDI MS/MS spectra for a pair of peptides at m/z 971.59 and 971.60. (C) Overlaid MS/MS spectra of (A) and (B).

MS/MS spectra. The labeling strategy is compatible with microbore LC-MALDI with the heated droplet interface and it should also be compatible with LC electrospray ionization MS. With microbore LC-MALDI, even using one-dimensional LC, low abundance proteins from the whey fraction of the bovine milk were detected. Phosphopeptides and milk protein variants are also detected with high confidence.

5.5 Literature Cited

(1) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

(2) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836-2842.

(3) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Mol. Cell. Proteomics* **2002**, *1*, 376-386.

(4) Qiu, Y.; Sousa, E. A.; Hewick, R. M.; Wang, J. H. *Anal. Chem.* **2002**, *74*, 4969-4979.

(5) Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem.* **2004**, *76*, 2748-2755.

(6) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., 3rd *Anal. Chem.* **2002**, *74*, 1650-1657.

(7) Chakraborty, A.; Regnier, F. E. *J. Chromatogr. A* **2002**, *949*, 173-184.

(8) Han, D. K.; Eng, J.; Zhou, H.; Aebersold, R. *Nat. Biotechnol.* **2001**, *19*, 946-951.

(9) Griffin, T. J.; Lock, C. M.; Li, X. J.; Patel, A.; Chervetsova, I.; Lee, H.; Wright, M. E.; Ranish, J. A.; Chen, S. S.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 867-874.

(10) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, *20*, 512-515.

(11) Spahr, C. S.; Susin, S. A.; Bures, E. J.; Robinson, J. H.; Davis, M. T.; McGinley, M. D.; Kroemer, G.; Patterson, S. D. *Electrophoresis* **2000**, *21*, 1635-1650.

(12) Wang, S.; Regnier, F. E. *J. Chromatogr. A* **2001**, *924*, 345-357.

(13) Gu, S.; Pan, S.; Bradbury, E. M.; Chen, X. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 1-7.

(14) Zhu, H.; Pan, S.; Gu, S.; Bradbury, E. M.; Chen, X. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2115-2123.

(15) Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 78-82.

(16) Oda, Y.; Huang, K.; Cross, F. R.; Cowburn, D.; Chait, B. T. *Proc. Natl. Acad. Sci. U S A* **1999**, *96*, 6591-6596.

(17) Jiang, H.; English, A. M. *J. Proteome Res.* **2002**, *1*, 345-350.

(18) Zappacosta, F.; Annan, R. S. *Anal. Chem.* **2004**, *76*, 6618-6627.

(19) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Anal. Chem.* **2003**, *75*, 6843-6852.

(20) Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. *Anal. Chem.* **2000**, *72*, 4047-4057.

(21) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163-170.

(22) Beardsley, R. L.; Reilly, J. P. *J. Proteome Res.* **2003**, *2*, 15-21.

(23) Johnson, K. L.; Muddiman, D. C. *J Am Soc Mass Spectrom* **2004**, *15*, 437-445.

(24) Staes, A.; Demol, H.; Van Damme, J.; Martens, L.; Vandekerckhove, J.; Gevaert, K. *J. Proteome Res.* **2004**, *3*, 786-791.

(25) Stewart, II; Thomson, T.; Figeys, D. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 2456-2465.

(26) Mirgorodskaya, O. A.; Kozmin, Y. P.; Titov, M. I.; Korner, R.; Sonksen, C. P.; Roepstorff, P. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1226-1232.

(27) Shui, W.; Liu, Y.; Fan, H.; Bao, H.; Liang, S.; Yang, P.; Chen, X. *J. Proteome Res.* **2005**, *4*, 83-90.

(28) Zhong, H.; Marcus, S. L.; Li, L. *J. Proteome Res.* **2004**, *3*, 1155-1163.

(29) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015-1024.

(30) Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* **1996**, *17*, 945-953.

(31) Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, S.; Zerial, M.; Wilm, M. *Proteomics* **2001**, *1*, 668-682.

(32) Goodlett, D. R.; Keller, A.; Watts, J. D.; Newitt, R.; Yi, E. C.; Purvine, S.; Eng, J. K.; von Haller, P.; Aebersold, R.; Kolker, E. *Rapid Commun. Mass Spectrom.* **2001**, *15*, 1214-1221.

(33) Hunt, D. F.; Yates, J. R., 3rd; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U S A* **1986**, *83*, 6233-6237.

(34) Keough, T.; Lacey, M. P.; Youngquist, R. S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2348-2356.

(35) Gu, S.; Pan, S.; Bradbury, E. M.; Chen, X. *Anal. Chem.* **2002**, *74*, 5774-5785.

(36) Hsu, J. L.; Huang, S. Y.; Shiea, J. T.; Huang, W. Y.; Chen, S. H. *J. Proteome Res.* **2005**, *4*, 101-108.

(37) Ji, C.; Li, L. *J. Proteome Res.* **2005**, *4*, 734-742.

(38) Ji, C.; Li, L. J.; Gebre, M.; Pasdar, M.; Li, L. *J. Proteome Res.* **2005**, *4*, 1419-1426.

(39) Fu, Qiang; Li, Lingjun "De novo sequencing of novel neuropeptides by a combination of N-terminal derivatization and nanoLC MS/MS", in *Proceedings of the American Society for Mass Spectrometry conference*, June 2005, San Antonio, TX.; Poster 294.

(40) Kimmel, J. R. *Methods Enzymol.* **1967**, *11*, 584-589.

(41) Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, *74*, 1884-1890.

(42) Bonetto, V.; Bergman, A. C.; Jornvall, H.; Sillard, R. *Anal. Chem.* **1997**, *69*, 1315-1319.

(43) Molloy, M. P.; Herbert, B. R.; Yan, J. X.; Williams, K. L.; Gooley, A. A. *Electrophoresis* **1997**, *18*, 1073-1078.

(44) Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087-1091.

(45) Zhang, B.; McDonald, C.; Li, L. *Anal. Chem.* **2004**, *76*, 992-1001.

(46) Beardsley, R. L.; Karty, J. A.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2147-2153.

(47) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2070-2073.

(48) Zhang, R.; Sioma, C. S.; Thompson, R. A.; Xiong, L.; Regnier, F. E. *Anal. Chem.* **2002**, *74*, 3662-3669.

(49) Zhang, R.; Regnier, F. E. *J. Proteome Res.* **2002**, *1*, 139-147.

(50) Zhang, R.; Sioma, C. S.; Wang, S.; Regnier, F. E. *Anal. Chem.* **2001**, *73*, 5142-5149.

(51) Holland, J. W.; Deeth, H. C.; Alewood, P. F. *Proteomics* **2004**, *4*, 743-752.

(52) Griffin, T. J.; Gygi, S. P.; Ideker, T.; Rist, B.; Eng, J.; Hood, L.; Aebersold, R. *Mol. Cell. Proteomics* **2002**, *1*, 323-333.

(53) Baeker, R.; Haebel, S.; Schlatterer, K.; Schlatterer, B. *Prostaglandins Other Lipid Mediators* **2002**, *67*, 75-88.

(54) Yamada, M.; Murakami, K.; Wallingford, J. C.; Yuki, Y. *Electrophoresis* **2002**, *23*, 1153-1160.

Chapter 6

Evaluation of Quantitative Reproducibility Using 2MEGA Labeling and LC-MALDI MS [a]

6.1 Introduction

Changes in protein expression profiles can provide critical information for the discovery of diagnostic or prognostic protein markers, the detection of new therapeutic targets, and the understanding of basic biological processes and mechanisms. Quantitative proteomics measures molecular physiology at the protein level and allows comparisons between samples by measuring relative changes in protein expression in response to external (pharmacological, environmental) or internal (genetic, pathological) perturbations [1-3]. An important consideration in the quantitative measurement of protein expression alteration is the accuracy as well as the reproducibility of the experiments. A quantitative assessment of the technical and biological variability is crucial to avoid erroneous inferences and conclusions.

Several techniques are available for measuring changes in protein expression. For more than 25 years, high-resolution two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE) has been the method of choice for quantitative proteomic profiling. Protein mixtures were initially separated by 2D PAGE followed by image analysis for relative protein quantification, and mass spectrometric analysis for identification. Several studies have been conducted to investigate the effects of technical variability of 2D PAGE in quantitative analysis [4-8]. Molly and coworkers suggested a coefficient of variation (CV) of 20-30% for technical variability in 2D PAGE [4]. Although the 2D GE approach has been very effective and consistent in the analysis of quantitative changes in protein expression, several technical disadvantages would limit the application of 2D PAGE for comprehensive quantitative proteomics studies. Those shortcomings include: (1) the procedure is laborious and difficult to automate; (2) it is difficult to analyze certain types of proteins with low or high molecular weight, very basic or acidic pI and high hydrophobicity; (3) a given spot might contain multiple proteins; and (4) the commonly used staining methods, such as silver-staining, may not be quantitative.

---

[a] Dr. N Zhang grew the cells and fractionated the membrane proteins.

In the past six years, non-gel based technologies for quantitative profiling have been rapidly developed since the first isotope coded affinity tag (ICAT) method was introduced by Aebersold's group [9]. These techniques have become powerful methods and have been widely used for high-throughput screening of protein expression alteration because of their potential ability to overcome the limitations of 2D PAGE-based methods for quantitative profiling. Even though many studies have been reported for quantitative proteome analysis using stable isotopic labeling, only a few studies addressed the range and the origins of variations introduced during the entire process of relative quantification of proteomes [7, 10]. To date, most of the stable-isotope-labeling quantitative proteomics strategies have been carried out using microcapillary (μ) LC-ESI-based methods, in which peptide identification precedes quantification, and identification and quantification are coupled. This has significant implications for quantitative proteomic analysis using stable isotope dilution. The main consequence is that numerous peptides identified are derived from proteins of unchanged abundance between the samples. In a typical comparative proteome study only 10–20% of the proteins between the two samples showed a difference in relative abundance [11, 12]. Therefore, most of the instrument time, data analysis, and operating time were wasted on redundantly analyzing peptides with little biological significance to the biosystem of interest. In addition, these constitutively represented peptides are repeatedly selected for MS/MS analysis, thus preventing the analysis of more informative peptides that are derived from proteins showing changes in abundance. To overcome the shortcomings from automated ESI methods, LC-MALDI based strategies were developed for quantitative proteome analysis and have been proved to be efficient compared to ESI analysis [13-16], because most of the effort (including time consumed for MS/MS data collecting and data analysis) can be focused on analyzing the peptide pairs resulting from differentially expressed proteins which have significant biological meanings.

In Chapter 5, an integrated strategy of 2MEGA differential isotope labeling with a microbore LC-MALDI interface for abundance ratio-dependent quantitative analysis has been reported [Ji #57]. Besides its data-dependent analysis, which is a common feature of all LC-MALDI-based approaches, this integrated strategy has some distinguishing features: (1) the isotopic labeling reaction is fast and has no side reactions; (2) the labeling reagent is inexpensive and commercially available; (3) the uniform 6 Da mass difference between peptide pairs simplifies the data analysis procedure; (4) it can be used to label peptides from any source; (5) the isotope effect is negligible and this can potentially increase the measurement accuracy of relative abundance ratio; (6) a microbore column with a large loading capacity, instead of a microcapillary column, is employed in the peptide mixture separation, therefore increase the detectability.

The key step of using the above integrated strategy for abundance ratio dependent quantitative proteome analysis is to define the abundance ratio threshold. A very stringent threshold will lose important information to peak list generation and further MS/MS analysis. However, a very low threshold will result in wasting efforts on analyzing data with little constructive meaning. In this study, the focus is to address the precision or reproducibility issue of the quantitative proteome analysis, i.e., how the technical variation affects the quantitative results. Quantitative variation can result from any individual experimental step, such as cell growth, protein extraction, enrichment of membrane proteins, protein sample workup, and MS analysis. It is very important to find out which step is the one that gives the greatest variation and should be focus during future analysis or experiment to improve precision. The results of this type of study would provide some guidance as to what peak ratio threshold should be used in quantitative proteome analysis. In this work, cultured leukemia (CCRF-CEM) cells were studied as a model to investigate the experimental variations in quantitative proteome analysis using 2MEGA LC-MALDI MS.

6.2 Experimental

6.2.1 Chemicals and Reagents

d(0),^{12}C-Formaldehyde (37% (w/w) in H_2O), *O*-methylisourea, sodium hydroxide, sodium bicarbonate, sodium cyanoborohydride, Leucine enkephalin (Leu-enk), bovine trypsin and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (Oakville, ON, Canada). d(2),^{13}C-formaldehyde (~20% (w/w) in deuterated water) was obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA). Acetonitrile was purchased from Fisher Scientific Canada (Edmonton, AB, Canada). Formic acid was obtained from Pierce. Water used in these experiments was obtained from a Milli-Q Plus purification system (Millipore, Bedford, MA). The MALDI matrix, 2,5-dihydroxybenzoic acid (DHB), was purchased from Aldrich (Milwaukee, WI).

6.2.2 Cell Growth

The CCRF-CEM cell line was originally obtained from Dr. W. T. Beck (formerly of St. Jude Children's Research Hospital, now at University of Illinois, Chicago, IL).

Cells were maintained between 3 × 10^5 and 7 × 10^5 cells per mL in a humidified incubator at 5% carbon dioxide in RPMI 1640 medium supplemented with 10% fetal calf serum (ATCC).

6.2.3 Membrane Protein Isolation and Tryptic Digestion

The Mem-PER membrane protein extraction kit (Pierce, Rockford, USA) was used for total membrane protein extraction. In brief, about 5 × 10^7 cells were washed with cold PBS, and pelleted in a 1.5 mL centrifuge tube. 150 µL of Reagent A was added to the cell pellet and was pipetted up and down to obtain a homogeneous cell suspension. After incubation for 10 min at room temperature with occasional vortexing, 450 µL Reagent B and C (1:2) was added to lysed cells and vortexed every 5 min with incubation on ice. Then the tube was centrifuged at 10 000 g for 3 min at 4 °C, and the supernatant was transferred to a new tube and incubated for 10 min at 37 °C. The phase separation was performed by centrifugation at 10 000 g for 2 min, and immediately the bottom layer, where the membrane fraction was located, was collected for further analysis.

Proteins in the collected membrane fraction were digested with trypsin. In brief, the solution was buffered with 1M $NaHCO_3$ (pH 8.5) to a final $NaHCO_3$ concentration of 50 mM. Trypsin was added to the protein solution at a ratio of 10:1, and $CaCl_2$ was added to the digestion solution to a final concentration of 2 mM to prevent trypsin autolysis.

6.2.4 2MEGA Labeling

Guanidination of lysine residues was performed as described previously [16, 18-20] with some modification. Trypsin in the 100 μL tryptic digest solution (about 1 μg/μL) was irreversibly inactivated by adding 10 μL 2 M sodium hydroxide. The ε-amino groups of all lysines were blocked by adding 40 μL 2 M *O*-methylisourea in 100 mM $NaHCO_3$, adjusting to pH 11 with 2 M sodium hydroxide and incubating the resulting mixture at 65 °C for 10 min. Then the reaction was stopped and the pH was adjusted to 8 by adding 10% TFA. Reductive amination with either d(0), ^{12}C-formaldehyde or d(2), ^{13}C-formaldehyde was also carried out as described previously [Hsu, 2003 #60;Hsu, 2005 #61;Ji #57;Ji, 2005 #14;Ji, 2005 #16] with some modifications. The guanidinated peptide solution was mixed with 15 μL 2 M sodium cyanoborohydride. The mixture was then vortexed and mixed with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde (4% (w/w) in water, 3 μL). The mixtures were vortexed and incubated at 37 °C for 1 h. If necessary, ammonium bicarbonate (1 M, 3 μL) was added to consume the excess formaldehyde.

6.2.5 Desalting Using Solid Phase Extraction Cartridge

The 2MEGA labeled peptide mixture solutions and the unlabeled half of the pooled tryptic digests from methanol-assisted and SDS-assisted digestion were desalted by SPE using bonded phase octadecyl (C-18) cartridges. Each cartridge was equilibrated with three bed volumes of acetonitrile and washed with three volumes of aqueous 0.1% TFA. The peptide mixture was applied to the cartridge and the cartridge was washed with three volumes of 0.1% TFA. Finally, the peptides were eluted initially with 500 µL of acetonitrile:H_2O:TFA (50:49.9:0.1, v/v/v) and then 1 mL acetonitrile:H_2O:TFA (75:24.9:0.1, v/v/v). The eluates were concentrated to 300 µL in a SpeedVac. The peptide mixture was stored at -20 °C.

6.2.6 Cation Exchange Chromatography

The nonionic detergent in the desalted peptide mixture was removed by strong cation exchange (SCX) chromatography on an Agilent 1100 HPLC system (Palo Alto, CA) using a 2.1 x 150 mm Hydrocell™ SP 1500 column (5 µm, Catalog No.: 24-34 SP, BioChrom Labs, Inc., Terre Haute, IN). The buffer solutions used were 20% v/v acetonitrile in 0.1% TFA (buffer A) and 20% v/v acetonitrile in 0.1% TFA, 1 M NaCl (buffer B). About 300 µL (~500 µg) proteins digest was loaded onto the SCX column and peptides were eluted with a linear gradient of 0-50% B in 6 min. The eluate was collected in a 2 mL vial based on the UV chromatography signal recorded 214 nm, and concentrated to ~150 µL in a SpeedVac.

6.2.7 Microbore LC-MALDI QqTOF Mass Spectrometric Analysis

About 100 µL (~330 µg) 2MEGA differential isotope labeled peptide mixture from the last step was separated by reversed-phase chromatography on a 1.0 × 150 mm

Vydac C_{18} column (5 µm particles with 300 Å pore size, Catalog No.: 218TP5115) at a flow-rate of 40 µL/min in an Agilent 1100 capillary HPLC equipped with an auto-sampler. Gradient elution was performed with solvent A (Milli-Q water, 0.1% TFA and 4% acetonitrile, v/v/v) and B (0.1% TFA in acetonitrile, v/v). The gradients were 0-5% B in 5 min, 5-40% B in 85 min, 40-90% B in 15 min. Before gradient elution, peptides in each SCX fraction were desalted by pumping with 100% solution A for 30 min. UV absorbance was recorded at 214 nm. Immediately after desalting, the HPLC eluate was directly collected in 1 min fractions onto a 100-well MALDI plate (Applied Biosystems, Concord, ON, Canada) using a home-built, heated droplet LC-MALDI interface [23]. After the fractionation was completed, the dried peptides in each well were redissolved and mixed with DHB matrix by the addition of 0.8 µL of 100 mg/mL DHB matrix in 50%ACN/50%water (v/v).

MALDI MS data were acquired on an Applied Biosystems/MDS-Sciex QSTAR Pulsar QqTOF instrument equipped with an orthogonal MALDI source employing a 337 nm nitrogen laser (Concord, ON, Canada) that has been previously described [13, 14]. The instrument was operated in positive ion mode and collision-induced dissociation (CID) of peptides was achieved with argon as collision gas. Spectra were acquired and processed using Sciex supporting software.

6.2.8 Peak Pair Detection and Abundance Ratio Calculation

All MALDI MS spectra were converted into text files. All the text files were then analyzed by ProST Data™ software (Efeckta Technologies Corporation, Steamboat Springs, CO) in a batch mode. The mass differences between peptide pairs were set at 6.032 Da with mass tolerance at 0.08 Da. The relative abundance ratio of each peptide pair was calculated using summed monoisotopic peak intensities of the peptide pair during the whole LC profiling. All the detected pairs were manually checked.

6.2.9 Nano-LC ESI QTOF Mass Spectrometric Analysis

The 2MEGA differential isotope labeled peptide mixture was also analyzed using a Q-Tof Premier Mass Spectrometer (Waters, Manchester, UK) equipped with a nanoACQUITY™ Ultra Performance LC system (Waters, Milford, MA, USA). In brief, 2 µL peptide solution from each SCX fraction was injected onto a 75µm × 100mm Atlantis™ dC18 column (Waters, Milford, MA, USA). Solvent A consisted of 0.1% formic acid in water; and solvent B consisted of 0.1% formic acid in acetonitrile. Peptides were separated using gradients of 5-30% solvent B in 80 min, 30-90% solvent B in 10 min and 90-5% solvent B in 10 min, and electrosprayed into a Q-Tof Premier Mass Spectrometer, fitted with a nanoLockSpray source, at a flow rate of 250 nL/min. Mass spectra were acquired from *m/z* 300 to 1600 for 1 s followed by 3 data dependent MS/MS scans from 50 to 1900 m/z for 1 s each. The collision energy used to perform MS/MS was varied according to the mass and charge state of the eluting peptide. Leu-enk (Lockmass) was infused at a rate of 250 nL/min and was acquired for 1 s every 2 min throughout the run. The exclusion list was generated based on Mascot searching results in which peptides with scores above the identity threshold were selected.

6.2.10 Protein Identification from MS/MS Data

Raw search data was lock mass corrected, de-isotoped and converted to peak list files by ProteinLynx Global Server 2.1.5 (Waters). Peptide sequences were identified via automated database searching of peak list files using the Mascot search program (Matrix Science, London, United Kingdom). Peak list files were searched twice; in one case with fixed modification setting as N-terminal dimethylation with d(0), ^{12}C-formaldehyde and guanidination of K (lysine), the other with fixed modification setting as N-terminal dimethylation with d(2), ^{13}C-formaldehyde and guanidination of K. Database searching was restricted to *Homo sapiens* in Swiss-Prot database (UniProtKB/Swiss-Prot Release 47.7 of 16-Aug-2005). The following search parameters were selected for all database searching: enzyme, trypsin; missed cleavages, 3; peptide tolerance, ±30 ppm; MS/MS tolerance: 0.2 Da; peptide charge, (1+, 2+ and 3+); variable modification, oxidation (M).

6.3 Results and Discussion

In this study, the membrane protein fraction enriched using the MEM-PER kit from cultured human leukemia cells (CCRF-CEM) was studied as a model system to investigate the reproducibility of the reported quantitative strategy in Chapter 5, based on 2MEGA differential isotope labeling microbore LC-MALDI QqTOF MS. A membrane fraction was chosen because about 70% of all drug targets are membrane proteins [24]. In addition, the analysis of membrane proteins by MS has been a challenge because of the hydrophobic nature of membrane proteins [25-28]. In practice, membrane protein enrichment was normally achieved using the buoyant separation method, which needs a large number of cells and many experimental steps. The Mem-PER kit, a commercial method for membrane protein extraction, has recently become available and membrane proteins can be enriched using less time and also fewer cells than with the buoyant method [29]. However, because this commercial kit involves the use of Triton X series detergents that are notorious for deteriorating MS signals [30-34], the enriched membrane proteins have been normally analyzed by the SDS-PAGE method where the Triton detergents are removed during electrophoresis [29]. To analyze the enriched membrane proteins the use of Triton series detergents was combined with solution-based analyses in which the enriched membrane proteins are analyzed by tryptic digestion, 2MEGA differential isotope labeling, chromatographic separation and LC-MALDI MS and LC-ESI MS/MS for analysis of the resulting peptides. Inter- and intra-experimental variation from cells grown under identical conditions were examined to gauge the reproducibility of preparation of cultured cells and membrane proteins, protein assay, tryptic digestion, 2MEGA labeling, RPLC separation and MALDI mass spectrometric detection of peptides. Three levels of possible variation were investigated using the workflow illustrated in Figure 6.1: variation level I, technical variation starting with the same membrane preparation divided into three portions and analyzed independently for protein digestion, isotope labeling, LC separation and MALDI mass spectrometric detection; variation level II, technical variation starting with three independent membrane preparations derived from the same batch of cells; variation level III, biological and technical variations determined from membrane proteins prepared from three individual batches of cells. For each level of variation, one of three samples was used as the control and combined with either of the other two samples to make sure each level of variation was studied with two replicates.

Table 6.1 shows the statistical results of quantitative reproducibility at three different levels. All the measured CVs (or relative standard deviation) are less than 15%, which is normally used as a quantitative threshold for bioassay. The variation measured at level I is technical variation, which comes from tryptic digestion, isotope labeling, and LC-MALDI QqTOF MS analysis. The variation measured at level II is also called technical variation containing the variation from level I and any variation associated with protein assays. The variation measured at level III includes technical and some biological variations. The biological variation can be calculated using the following equation:

$$s_{total}^2 = s_{technical}^2 + s_{biological}^2$$

where s_{total} is the standard deviation calculated at level III and $s_{technical}$ is the standard deviation calculated at level II. Because the data set in each level has duplicates, the pooled standard deviation from both levels (II and III) was used to calculate biological

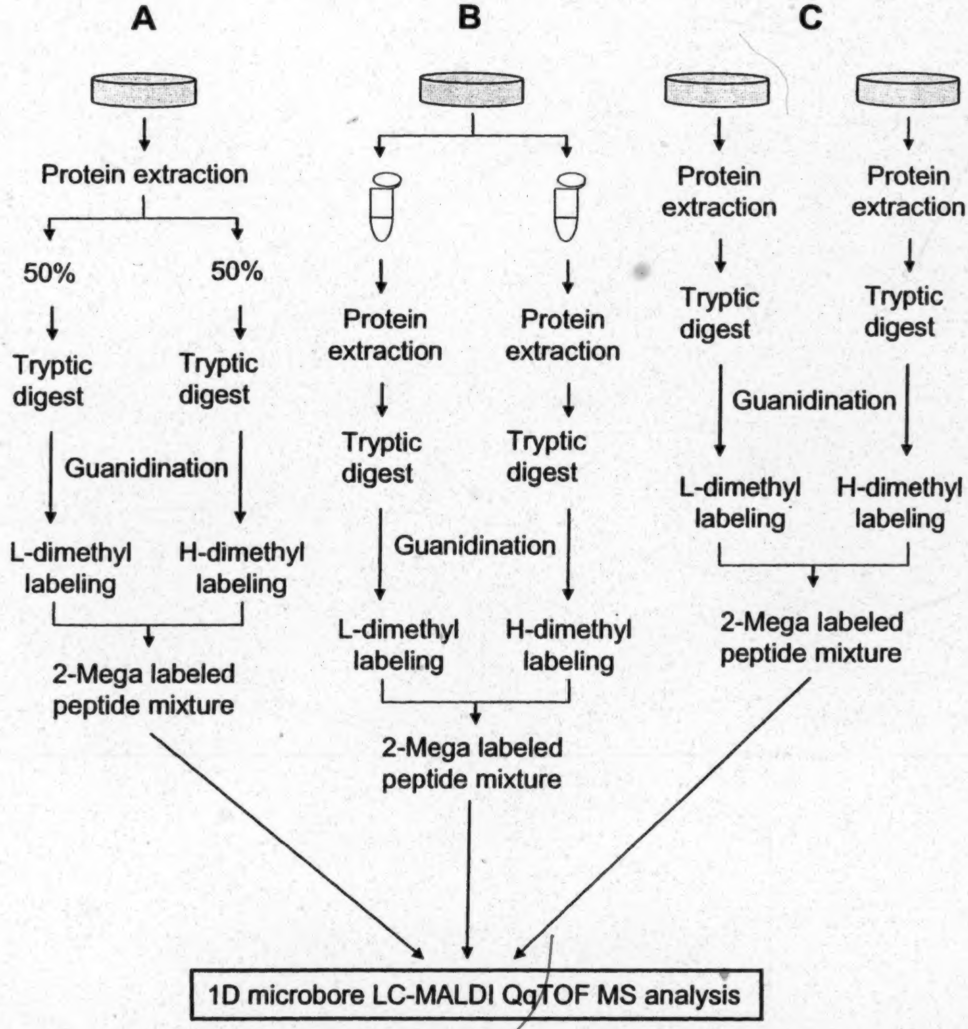
A
B
C
Protein extraction
50%
50%
Tryptic digest
Tryptic digest
Guanidination
L-dimethyl labeling
H-dimethyl labeling
2-Mega labeled peptide mixture
Protein extraction
Protein extraction
Tryptic digest
Tryptic digest
Guanidination
L-dimethyl labeling
H-dimethyl labeling
2-Mega labeled peptide mixture
Protein extraction
Protein extraction
Tryptic digest
Tryptic digest
Guanidination
L-dimethyl labeling
H-dimethyl labeling
2-Mega labeled peptide mixture
1D microbore LC-MALDI QqTOF MS analysis

Figure 6.1 Workflow for investigation of quantification variations using integrated strategy of N-terminal differential dimethyl labeling and microbore LC-MALDI QqTOF MS. A flowchart for measuring the quantitative reproducibility starting at (A) post proteins level. (B) membrane protein preparation level. (C) cell growth level.

Table 6.1 Summary of statistical analysis results of quantitative reproducibility at three different levels.

	# of peptide pairs	Average	Variation range	SD	pooled SD	CV	# of peptide >1.5 or <0.67	# of peptide >2 or <0.5
				Level I				
A	441	1.01	0.66-1.31	0.097	0.110	9.7	1	0
B	568	1.01	0.67-1.95	0.118		11.7	5	0
				Level II				
A	448	1.16	0.69-1.96	0.166	0.132	14.3	4	0
B	438	1.03	0.73-1.36	0.085		8.3	0	0
				Level III				
A	655	1.03	0.68-1.42	0.141	0.146	13.7	0	0
B	576	1.05	0.71-1.49	0.151		14.4	0	0

Note: SD: standard deviation; CV: coefficient of variation

variation. The calculated $s_{\text{biological}}^2$ is 0.0038, which is less than $s_{\text{technical}}^2$ (0.0174). Therefore, in the future, more attention should be paid to the sample handling, such as protein extraction, protein assay, tryptic digestion, labeling and microbore LC MALDI MS analysis, to reduce the technical variation. In addition, the calculated ratio for all detected peptide pairs with S/N greater than 3 in one measurement at level II is 1.16, slightly over the theoretical ratio of 1.00, indicating that normalization is required to reduce the measurement error from protein assays. The measured relative abundance ratios of all peptide pairs, with signal-to-noise ratio greater 3, in each experiment are well below 2-fold (see Figures 6.2 - 6.4 and Table 6.1), indicating that a relative abundance ratio of 2 could be set as a stringent threshold with 100% confidence, to quantify and identify differentially expressed proteins between two cultured cell lines with specific phenotypic differences. In practice, a smaller threshold of an abundance ratio of 1.5 could also be used for biological study reassessment, with an understanding that a very small number of peptide pairs may be falsely identified. The false identification rate in this case would be less than 1%, as indicated in Table 6.1.

In this study, the efficiency of enrichment of membrane proteins using the Mem-PER kit was also studied. Table 6.2 lists all proteins, identified from one membrane fraction prepared by using the Mem-PER kit, with their subcellular locations included. In total, 127 proteins, including 35 cytoplasmic proteins, 21 membrane or

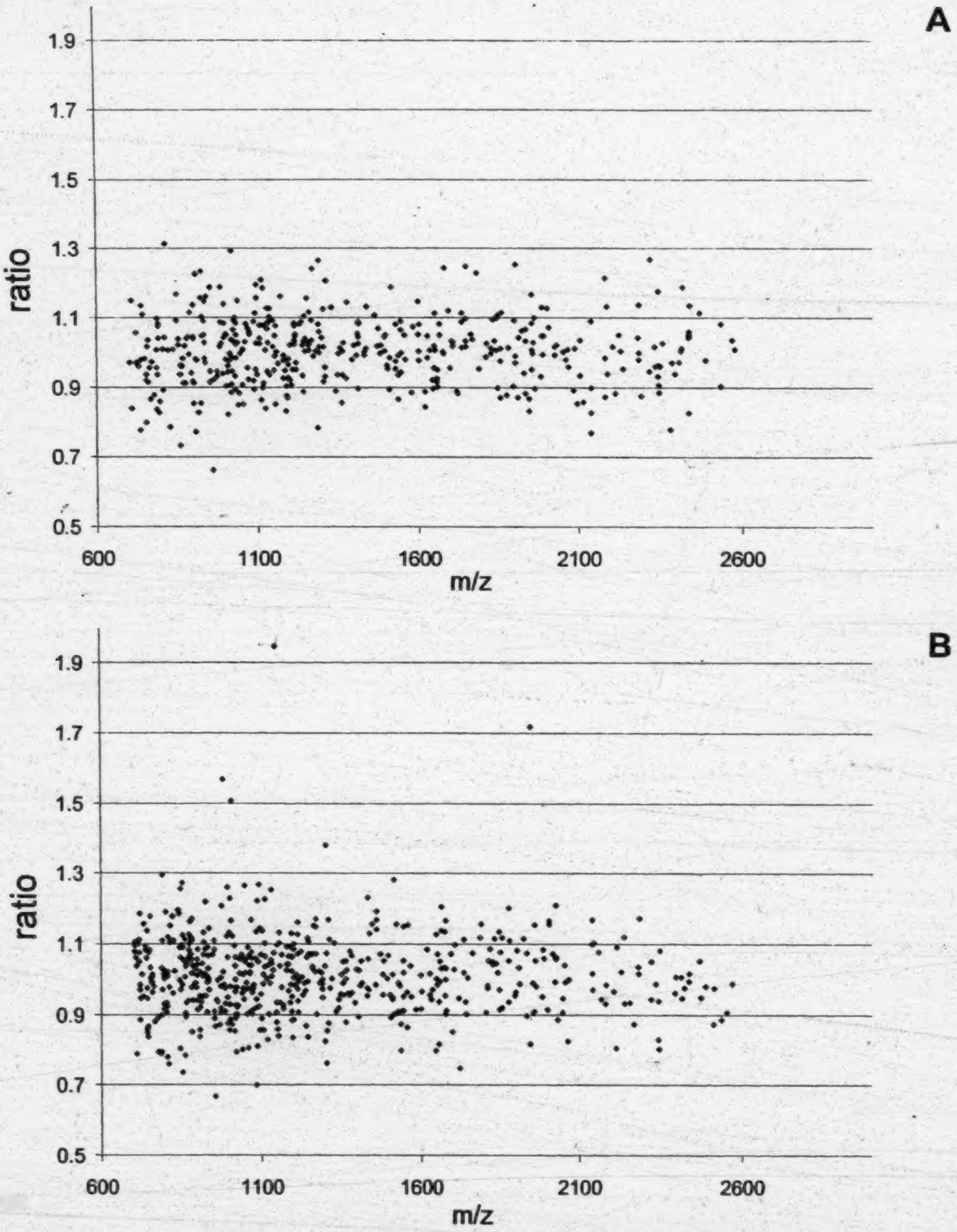

Figure 6.2 The measured ratios of all peptide pairs shown in MALDI MS spectra plotted against the protonated peptide masses of d(0), ^{12}C-formaldehyde labeled peptides from quantitative reproducibility analysis starting at post proteins level. (A) and (B) are from paralleled experiments.

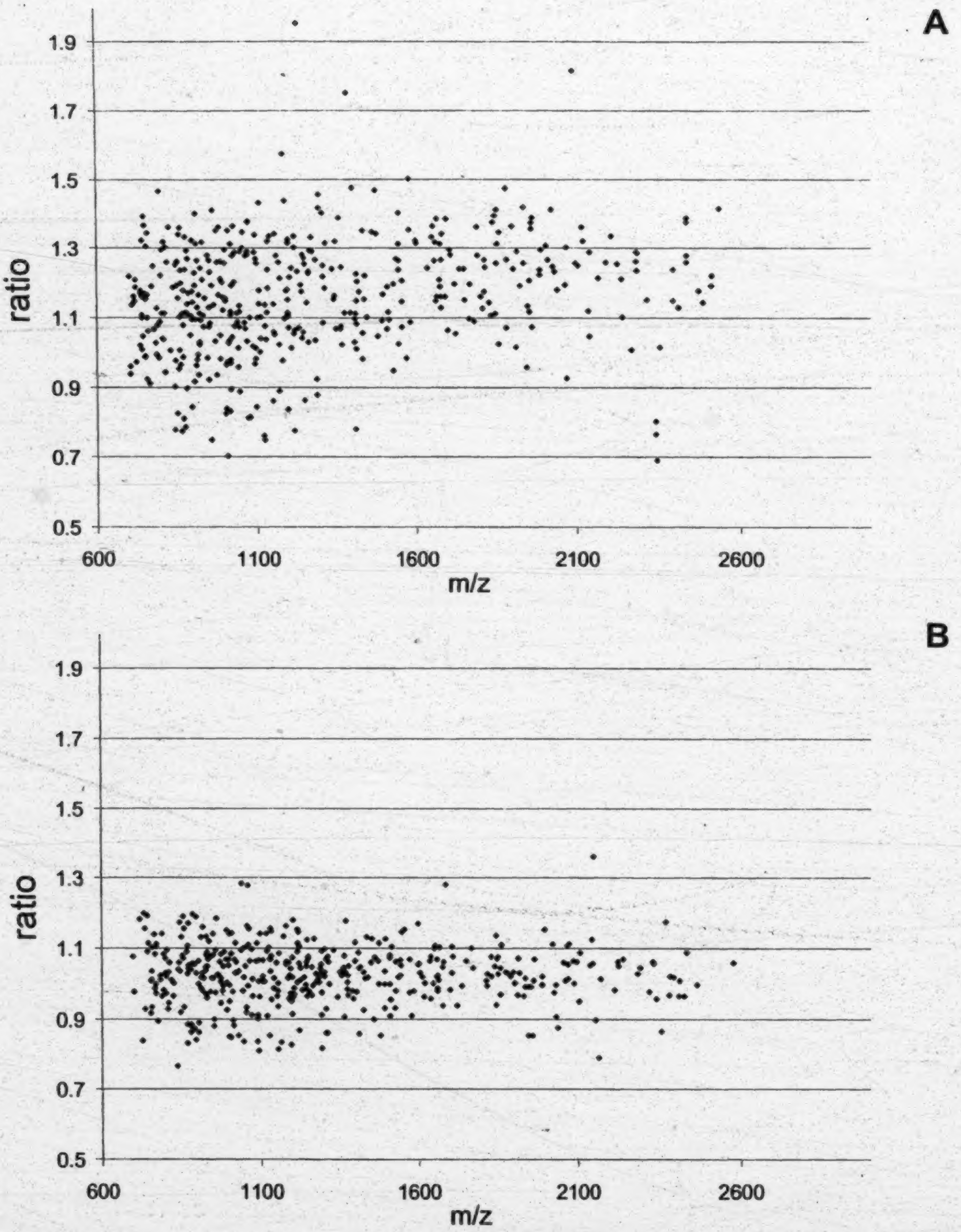

Figure 6.3 The measured ratios of all peptide pairs shown in MALDI MS spectra plotted against the protonated peptide masses of d(0), ^{12}C-formaldehyde labeled peptides from quantitative reproducibility analysis starting at membrane protein preparation level. (A) and (B) were plotted from duplicated experiments.

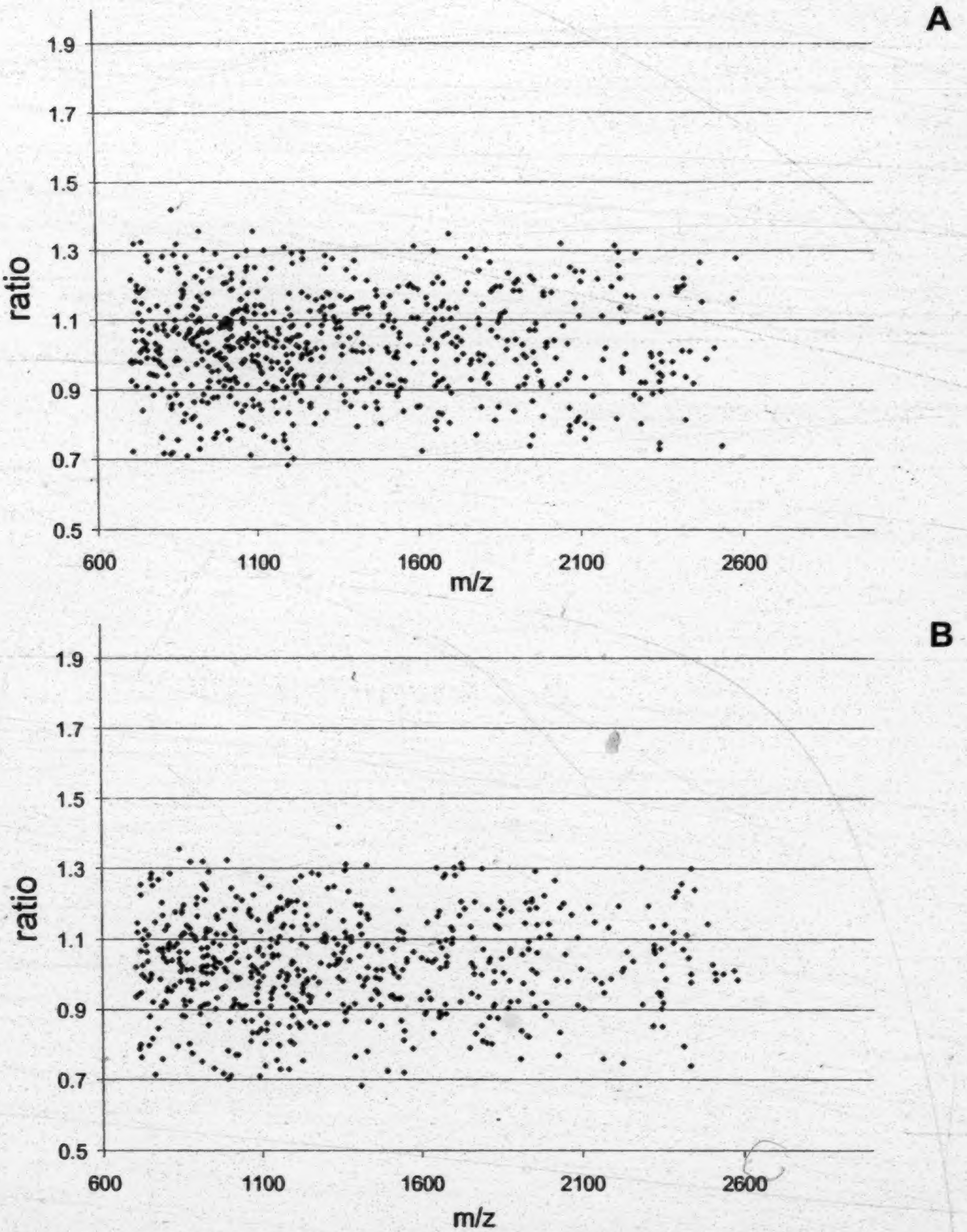

Figure 6.4 The measured ratios of all peptide pairs shown in MALDI MS spectra plotted against the protonated peptide masses of d(0), ^{12}C-formaldehyde labeled peptides from quantitative reproducibility analysis starting at cell growth level. (A) and (B) were plotted from duplicated experiments.

membrane-associated proteins, 24 nuclear proteins and 38 proteins with unknown subcellular location, were identified. Only 16.5% of total identified proteins are membrane or membrane-associated proteins, indicating that membrane protein enrichment by using MEM-PER kit is not efficient. Therefore, other enrichment methods, such as sucrose gradient and carbonate fractionation, should be investigated in the future. Using other enrichment methods may introduce a different degree of experimental variation.

6.4 Conclusions

The technical and biological variations of quantitative proteome analysis using the integrated strategy of 2MEGA differential isotope labeling and microbore LC-MALDI QqTOF MS were investigated. All the measured CVs are well below 15%, the value that is normally employed as a threshold in bioanalysis, indicating that the developed quantitative proteome analysis strategy has very good reproducibility. It was found that the biological variation originating from cell growth is less than technical variation from the post protein level, including protein assay, tryptic digestion, labeling, and microbore LC-MALDI mass spectrometric detection. The results revealed that for further biological quantitative analysis, more experimental attention should be paid to the analytical laboratory procedures than to the biological laboratory. It was demonstrated that a relative abundance ratio of greater than 2-fold can be set as a very stringent threshold to quantify and identify differentially expressed proteins between two cultured cell lines. In practice, a 1.5-fold threshold can be employed to provide a more detailed proteome profile to avoid loss of some biologically interesting information when the integrated strategy is used for comparative proteome analysis. In addition, the low percentage of membrane proteins identified in the membrane protein fraction enriched by the commercially available MEM-PER kit indicates the inefficiency of this enrichment method. Therefore, other membrane enrichment methods need to be investigated for membrane proteome analysis.

6.5 Literature Cited

(1) Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. *Science* **2001**, *292*, 929-934.

(2) Gygi, S. P.; Rist, B.; Aebersold, R. *Curr. Opin. Biotechnol.* **2000**, *11*, 396-401.

(3) Anderson, N. L.; Anderson, N. G. *Electrophoresis* **1998**, *19*, 1853-1861.

(4) Molloy, M. P.; Brzezinski, E. E.; Hang, J.; McDowell, M. T.; VanBogelen, R. A. *Proteomics* **2003**, *3*, 1912-1919.

(5) Mahon, P.; Dupree, P. *Electrophoresis* **2001**, *22*, 2075-2085.

(6) Voss, T.; Haberl, P. *Electrophoresis* **2000**, *21*, 3345-3350.

(7) Choe, L. H.; Aggarwal, K.; Franck, Z.; Lee, K. H. *Electrophoresis* **2005**, *26*, 2437-2449.

(8) Choe, L. H.; Lee, K. H. *Electrophoresis* **2003**, *24*, 3500-3507.

(9) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat.*

Biotechnol. **1999**, *17*, 994-999.

(10) Molloy, M. P.; Donohoe, S.; Brzezinski, E. E.; Kilby, G. W.; Stevenson, T. I.; Baker, J. D.; Goodlett, D. R.; Gage, D. A. *Proteomics* **2005**, *5*, 1204-1208.

(11) Li, J.; Steen, H.; Gygi, S. P. *Mol. Cell. Proteomics* **2003**, *2*, 1198-1204.

(12) Somiari, R. I.; Sullivan, A.; Russell, S.; Somiari, S.; Hu, H.; Jordan, R.; George, A.; Katenhusen, R.; Buchowiecka, A.; Arciero, C.; Brzeski, H.; Hooke, J.; Shriver, C. *Proteomics* **2003**, *3*, 1863-1873.

(13) Ji, C.; Li, L.; Gebre, M.; Pasdar, M. *J. Proteome Res.* **2005**, *4*, 1419-1426.

(14) Ji, C.; Li, L. *J. Proteome Res.* **2005**, *4*, 734-742.

(15) Griffin, T. J.; Lock, C. M.; Li, X. J.; Patel, A.; Chervetsova, I.; Lee, H.; Wright, M. E.; Ranish, J. A.; Chen, S. S.; Aebersold, R. *Anal. Chem.* **2003**, *75*, 867-874.

(16) Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem.* **2004**, *76*, 2748-2755.

(17) Ji, C.; Guo, N.; Li, L. *J. Proteome Res.* **2005**, in press.

(18) Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, *74*, 1884-1890.

(19) Kimmel, J. R. *Methods Enzymol.* **1967**, *11*, 584-589.

(20) Zappacosta, F.; Annan, R. S. *Anal. Chem.* **2004**, *76*, 6618-6627.

(21) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Anal. Chem.* **2003**, *75*, 6843-6852.

(22) Hsu, J. L.; Huang, S. Y.; Shiea, J. T.; Huang, W. Y.; Chen, S. H. *J. Proteome Res.* **2005**, *4*, 101-108.

(23) Zhang, B.; McDonald, C.; Li, L. *Anal. Chem.* **2004**, *76*, 992-1001.

(24) Hopkins, A. L.; Groom, C. R. *Nat Rev Drug Discov* **2002**, *1*, 727-730.

(25) Wu, C. C.; Yates, J. R., 3rd *Nat. Biotechnol.* **2003**, *21*, 262-267.

(26) Santoni, V.; Molloy, M.; Rabilloud, T. *Electrophoresis* **2000**, *21*, 1054-1070.

(27) le Maire, M.; Champeil, P.; Moller, J. V. *Biochim Biophys Acta* **2000**, *1508*, 86-111.

(28) Almgren, M. *Biochim Biophys Acta* **2000**, *1508*, 146-163.

(29) Qoronfleh, M. W.; Benton, B.; Ignacio, R.; Kaboord, B. *J Biomed Biotechnol* **2003**, *2003*, 249-255.

(30) Shore, L. J.; Odell, G. B.; Fenselau, C. *Biochem Pharmacol* **1995**, *49*, 181-186.

(31) Schnaible, V.; Michels, J.; Zeth, K.; Freigang, J.; Welte, W.; Buhler, S.; Glocker, M. O.; Przybylski, M. *International Journal of Mass Spectrometry* **1997**, *169*, 165-177.

(32) Rosinke, B.; Strupat, K.; Hillenkamp, F.; Rosenbusch, J.; Dencher, N.; Kruger, U.; Galla, H. J. *Journal of Mass Spectrometry* **1995**, *30*, 1462-1468.

(33) Ogorzalek Loo, R. R.; Dales, N.; Andrews, P. C. *Protein Science* **1994**, *3*, 1975-1983.

(34) Zhang, N.; Li, L. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 889-896.

Table 6.1 Identification of proteins in membrane protein fraction enriched using MEM-PER kit.

#	Protein Name	AccessID	Unique Peptide Sequence*	Subcellular Location
1	Phosphoglycerate kinase 1	P00558	DVLFLK ELNYFAK LGDVYVNDAFGTAHR VSHVSTGGGASLELLEGK ALESPERPFLAILGGAK ITLPVDFVTADKFDENAK	Cytoplasmic
2	Glyceraldehyde-3-phosphate dehydrogenase	P04406	VVDLMAHMASK VVDLMAHMASKE LISWYDNEFGYSNR VIISAPSADAPMFVMGVNHEK	Cytoplasmic
3	Heat shock protein HSP 90-alpha	P07900	FYEQFSK ELHINLIPNK RAPFDLFENR HIYYITGETK HFSVEGQLEFR SLTNDWEDHLAVK	Cytoplasmic
4	Heat shock protein HSP 90-beta	P08238	FYEAFSK RAPFDLFENK HFSVEGQLEFR SLTNDWEDHLAVK YESLTDPSKLDSGK HLEINPDHPIVETLR	Cytoplasmic
5	Alpha enolase	P06733	YDLDFK SGKYDLDFK IGAEVYHNLK YISPDQLADLYK LMIEMDGTENKSK+Oxidation(M)	Cytoplasmic
6	Heat shock cognate 71 kDa protein	P11142	MVNHFIAEFK RFDDAVVQSDMK	Translocates rapidly from the cytoplasm to the nuclei, and especially to the nucleoli, upon heat shock

			SQIHDIVLVGGSTR	
			IINEPTAAAIAYGLDKK	
7	ATP synthase alpha chain	P25705	GYLDKLEPSK	Mitochondrial inner membrane
			HALIIYDDLSK	
			EAYPGDVFYLHSR	
8	ATP synthase beta chain	P06576	IMDPNIVGSEHYDVAR	Mitochondrial
			FLSQPFQVAEVFTGHMGK	
			EGNDLYHEMIESGVINLK	
9	Rho GDP-dissociation inhibitor 2	P52566	ELQEMDKDDESLIK	Cytoplasmic
			APEPHVEEDDDDELDSK	
10	Fructose-bisphosphate aldolase A	P04075	ADDGRPFPQVIK	
			FSHEEIAMATVTALR	
			IGEHTPSALAIMENANVLAR	
11	Nucleosome assembly protein 1-like 1	P55209	FYEEVHDLER	Nuclear
			NVDLLSDMVQEHDEPILK	
12	Triosephosphate isomerase	P60174	FFVGGNWK	
			HVFGESDELIGQK	
			KQSLGELIGTLNAAK	
13	Rab GDP dissociation inhibitor beta	P50395	DWNVDLIPK	
			FKIPGSPPESMGR	
			MTGSEFDFEEMKR	
14	Putative nucleoside diphosphate kinase	O60361	DRPFFPGLVK	
			FKPEELVDYK	
			VMLGETNPADSKPGTIR	
15	Phosphoglycerate mutase 1	P18669	FSGWYDADLSPAGHEEAK	
16	Pyruvate kinase, isozymes M1/M2	P14618	LNFSHGTHEYHAETIK	
			TATESFASDPILYRPVAVALDTK	
17	High mobility group protein 1	P09429	IKGEHPGLSIGDVAK	Nuclear
			KHPDASVNFSEFSK	
18	60 kDa heat shock protein	P10809	ISSIQSIVPALEIANAHR	Mitochondrial matrix
19	Calnexin	P27824	TPELNLDQFHDK	Type I membrane protein
			KIPNPDFFEDLEPFR	
20	Stress-70 protein	P38646	VEAVNMAEGIIHDTETK	Mitochondrial

21	Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	DYFEEYGK	Nuclear
			GFGFVTFDDHDPVDK	
22	Glutathione S-transferase P	P09211	TLGLYGK	
			ALPGQLKPFETLLSQNQGGK	
23	Glucose-6-phosphate isomerase	P06744	EWFLQAAK	Cytoplasmic
			VWYVSNIDGTHIAK	
24	Nucleoside diphosphate kinase A	P15531	DRPFFAGLVK	Nuclear and cytoplasmic
			VMLGETNPADSKPGTIR	
25	T-complex protein 1, delta subunit	P50991	DALSDLALHFLNK	Cytoplasmic
26	Sodium/potassium-transporting ATPase beta-3 chain	P54709	IIGLKPEGVPR	Type II membrane protein
			SDPTSYAGYIEDLKK	
27	Translationally controlled tumor protein	P13693	DLISHDEMFSDIYK	Cytoplasmic
28	Tubulin-specific chaperone A	O75347	ILENEKDLEEAEEYK	
29	Calmodulin	P62158	EAFSLFDKDGDGTITTK	
30	Ubiquinol-cytochrome-c reductase complex core protein I	P31930	ADLTEYLSTHYK	Mitochondrial inner membrane
31	Vimentin	P08670	ETNLDSLPLVDTHSK	
32	Peptidyl-prolyl cis-trans isomerase A	P62937	TEWLDGK	Cytoplasmic
			VSFELFADKVPK	
33	Inosine-5'-monophosphate dehydrogenase 2	P12268	FGVPVIADGGIQNVGHIAK	
34	Nuclear pore complex protein Nup155	O75694	SSTAISSIAADGEFLHELEEK	Nuclear pore Nucleoplasmic and cytoplasmic peripheral membrane protein
35	ATP-dependent DNA helicase II	P13010	ANPQVGVAFPHIK	Nuclear
36	Heat shock-related 70 kDa protein 2	P54652	IINEPTAAAIAYGLDKK	
37	Import inner membrane translocase subunit TIM44	O43615	VFEPNEEALGVVLHK	Mitochondrial inner membrane
38	Alanyl-tRNA synthetase	P49588	FIDFFK	Cytoplasmic (Potential)
39	Radixin	P35241	IGFPWSEIR	Highly concentrated in the undercoat of the cell-to-cell adherens junction and the cleavage furrow in the interphase and mitotic phase, respectively

40	Far upstream element binding protein 1	Q96AE4	AWEEYYK IQFKPDDGTTPER	Nuclear (Probable)
41	Rho GDP-dissociation inhibitor 1	P52565	SIQEIQELDKDDESLR	Cytoplasmic
42	Barrier-to-autointegration factor	O75531	DFVAEPMGEKPVGSLAGIGEVLGK	Nuclear
43	10 kDa heat shock protein	P61604	FLPLFDR KFLPLFDR	Mitochondrial matrix
44	Lamin B1	P20700	DAALATALGDKK	Nucleoplasmic side of the inner nuclear membrane
45	60S ribosomal protein L10a	P62906	AVDIPHMDIEALK	
46	L-lactate dehydrogenase B chain	P07195	LKDDEVAQLK	Cytoplasmic
47	Tropomyosin alpha 4 chain	P67936	YSEKEDKYEEEIK	
48	Malate dehydrogenase, cytoplasmic	P40925	DVIATDKEDVAFK EVGVYEALKDDSWLK	Cytoplasmic
49	Phosphoserine aminotransferase	Q9Y617	GVGISVLEMSHR	
50	Malate dehydrogenase	P40926	VNVPVIGGHAGK	Mitochondrial matrix
51	Phosphatidylinositol transfer protein beta isoform	P48739	MIAPEGSLVFHEK	Cytoplasmic, and golgi
52	Eukaryotic translation initiation factor 5A	P63241	EDLRLPEGDLGK	
53	L-plastin	P13796	ISFDEFIK	Cytoplasmic
54	Peroxiredoxin 2	P32119	KEGGLGPLNIPLLADVTR	Cytoplasmic
55	Copine-1	Q99829	SDPFLEFFR	
56	Acyl-CoA-binding protein	P07108	WDAWNELK	
57	Cytochrome c	P99999	TGPNLHGLFGR	Mitochondrial matrix
58	Activator of 90 kDa heat shock protein ATPase homolog 1	O95433	VFTTQELVQAFTHAPATLEADR	Cytosolic
59	Complement component 1, Q subcomponent binding protein	Q07021	AFVDFLSDEIKEER	Mitochondrial matrix
60	Proteasome activator complex subunit 1	Q06323	LEGFHTQISK	
61	Voltage-dependent anion-selective channel protein 1	P21796	GYGFGLIK	Outer membrane of mitochondria and plasma membrane
62	Annexin A1	P04083	AAYLQETGKPLDETLK	
63	4F2 cell-surface antigen heavy chain	P08195	IGDLQAFQGHGAGNLAGLK	Type II membrane protein
64	PEST-containing nuclear protein	Q8WW12	SAEEEAADLPTKPTK	Nuclear

65	Calcyclin-binding protein	Q9HB71	WDYLTQVEK	Nuclear and cytoplasmic
66	Protein disulfide-isomerase	P07237	ILEFFGLK	Endoplasmic reticulum lumen
67	Elongation factor 2	P13639	EDLYLKPIQR	Cytoplasmic
68	Flap endonuclease-1	P39748	LDDFFK	Nuclear
69	60S ribosomal protein L12	P30050	PPKFDPNEIK	
70	Importin-alpha re-exporter	P55060	GSNTIASAAADKIPGLLGVFQK	Nuclear and cytoplasmic
71	Proliferating cell nuclear antigen	P12004	YLNFFTK	Nuclear
72	Cytochrome c oxidase polypeptide Vic	P09669	AYADFYR	Mitochondrial inner membrane
73	Ubiquitin-conjugating enzyme E2 L3	P68036	IEINFPAEYPFKPPK	
74	Transketolase	P29401	VLDPFTIKPLDR	
75	Neutral alpha-glucosidase AB	Q14697	YFTWDPSR	Endoplasmic reticulum and Golgi
76	Heterogeneous nuclear ribonucleoprotein M	P52272	NLPFDFTWK	Nuclear
77	Solute carrier family 2, facilitated glucose transporter member 5	P22732	VSEVYPEKEELK	Integral membrane protein
78	Ras-related C3 botulinum toxin substrate 1	P63000	WYPEVR	Inner surface of plasma membrane
79	Ras-related C3 botulinum toxin substrate 2	P15153	WFPEVR	Cytoplasmic; membrane-associated when activated
80	ADP/ATP translocase 2	P05141	EQGVLSFWR	Integral membrane protein
81	Elongation factor 1-gamma	P26641	VLSAPPHFHFGQTNR	
82	ATP synthase D chain	O75947	YTAQVDAEEKEDVK	
83	40S ribosomal protein S11	P62280	EAIEGTYIDKK	
84	Protein disulfide-isomerase A3	P30101	LSKDPNIVIAK	Endoplasmic reticulum lumen
85	Transgelin-2	P37802	DDGLFSGDPNWFPK	
86	L-lactate dehydrogenase A chain	P00338	NVNIFK	Cytoplasmic
87	Calreticulin	P27797	IKDPDASKPEDWDER	Endoplasmic reticulum lumen
88	Actin, cytoplasmic 2	P63261	DLTDYLMK	Cytoplasmic
		P60709	AVFPSIVGRPR	
			IWHHTFYNELR	
			VAPEEHPVLLTEAPLNPK	
			KDLYANTVLSGGTTMYPGIADR	
89	Transmembrane protease, serine 13	Q9BYE2	NKPGVYTK	Type II membrane protein
		P07477		
90	Tubulin alpha-1 chain	P68366	FDLMYAK	

		P68363	QLFHPEQLITGK	
			NLDIERPTYTNLNR	
			IHFPLATYAPVISAEK	
91	Actin, alpha cardiac	P68032	DLTDYLMK	Cytoplasmic
		P68133	AVFPSIVGRPR	
			IWHHTFYNELR	
92	Tubulin beta-2 chain	P07437	LHFFMPGFAPLTSR	
		P68371	GHYTEGAELVDSVLDVVR	
93	ATP-dependent helicase DDX39	O00148	DFLLKPELLR	Nuclear (By similarity)
		Q13838	GSYVSIHSSGFR	
94	Nuclease sensitive element binding protein 1	P67809	EDVFVHQTAIK	Nuclear
		Q9Y2T7		
		P16989		
95	Ras-related protein Rab-11B	Q15907	DDEYDYLFK	
		P62491		
96	TATA-binding protein associated factor 2N	Q92804	AAIDWFDGK	Nuclear (Potential)
		P35637		
97	Nucleoside diphosphate kinase B	P22392	DRPFFPGLVK	Nuclear and cytoplasmic
			TFIAIKPDGVQR	
			VMLGETNPADSKPGTIR	
98	Moesin	P26038	IGFPWSEIR	Cytoplasmic
			KTQEQLALEMAELTAR	
99	T-complex protein 1, theta subunit	P50990	VADMALHYANK+Oxidation(M)	Cytoplasmic
			HFSGLEEAVYR	
100	Mannosyl-oligosaccharide glucosidase	Q13724	LAGSLLTQALESHAEGFR	Type II membrane protein
101	Eukaryotic initiation factor 4A-I	P60842	LQMEAPHIIVGTPGR	
102	WD-repeat protein 1	O75083	NIDNPALADIYTEHAHQVVVAK	
103	40S ribosomal protein S20	P60866	LIDLHSPSEIVK	
104	Transferrin receptor protein 1	P02786	GFVEPDHYVVVGAQR	Type II membrane protein
105	Elongation factor 1-alpha 1	P68104	YYVTIIDAPGHR	Cytoplasmic
106	Ubiquitin-activating enzyme E1	P22314	DEFEGLFK	
107	Adenine phosphoribosyltransferase	P07741	DISPVLKDPASFR	Cytoplasmic
108	Isocitrate dehydrogenase	P48735	LNEHFLNTTDFLDTIK	Mitochondrial

109	Integrin beta-2	P05107	SAVGELSEDSSNVVHLIK	Type I membrane protein
110	Mago nashi protein homolog	P61326	IIDDSEITKEDDALWPPPDR	Nuclear (By similarity)
111	Peptidyl-prolyl cis-trans isomerase B	P23284	VIFGLFGK	Endoplasmic reticulum lumen
112	40S ribosomal protein S2	P15880	SPYQEFTDHLVK	
113	Heterogeneous nuclear ribonucleoprotein A1	P09651	DYFEQYGK	Nuclear
114	Ubiquitin thiolesterase protein OTUB1	Q96FW1	FFEHFIEGGR	
115	Endoplasmin	P14625	EFEPLLNWMK	Endoplasmic reticulum lumen
116	Adenosylhomocysteinase	P23526	KLDEAVAEAHLGK	Cytoplasmic
117	Microtubule-associated protein RP/EB family member 2	Q15555	DFYFGK	Associated with the microtubule network
118	ATP-binding cassette sub-family E member 1	P61221	AIIKPQYVDQIPK	Cytoplasmic and mitochondrial
119	Exportin-1	O14980	NVDILKDPETVK	Nuclear and cytoplasmic
120	Voltage-dependent anion-selective channel protein 2	P45880	GFGFGLVK	Outer mitochondrial membrane
121	Leucine zipper-EF-hand containing transmembrane protein 1	O95202	DFSVFFQK	Mitochondrial membrane protein
122	Septin-6	Q14141	TVPLAGHVGFDSLPDQLVNK	
123	Bifunctional purine biosynthesis protein PURH	P31939	ALFEEVPELLTEAEKK	
124	Cell division control protein 2 homolog	P06493	DLKPQNLLIDDK	Nuclear (By similarity)
125	Fatty acid-binding protein	Q01469	GFDEYMK	Cytoplasmic
126	Sodium channel beta-3 subunit	Q9NY72	LRCISCMK	Type I membrane protein
127	Cullin-associated NEDD8-dissociated protein 1	Q86VP6	DLLDTVLPHLYNETK	Nuclear (By similarity)

* The peptide sequences were identified only if the matched scores are above the MASCOT identity threshold.

Chapter 7

Large-Scale Evaluation of the Effect of 2MEGA Labeling for Membrane Proteome Analysis Using Nano-LC ESI QTOF [a]

7.1 Introduction

Mass spectrometry (MS)-based peptide sequencing has been accepted as one of the most reliable techniques for the identification of proteins and their post-translational modifications. Recent advancements in instrumentation [1-4], database searching engines [5-7], and high performance separation techniques [8-11] have led to an emergence of high-throughput approaches designed to identify thousands of peptides from a variety of biologically complex protein mixtures. The rapid and sensitive identification of thousands of peptides by 'shotgun proteomics' [12] strategies, the direct analysis of complex peptide mixtures derived from proteolytic digestion of heterogeneous mixtures of proteins by rapidly generating a global profile of the protein complement within the mixture, has become widespread. In these procedures, complex protein mixtures are typically digested by the enzyme trypsin to produce extremely complex peptide mixtures. The peptide mixtures are then subjected to extensive separations, such as strong cation exchange (SCX) chromatography, coupled online or offline with reversed-phase (RP) capillary liquid chromatography (LC). Peptides eluting from the RP LC column are analyzed by electrospray ionization (ESI) MS and selected ions are subjected to fragmentation by collision induced dissociation to produce MS/MS spectra. Peptide identifications are usually made by comparing the experimental MS/MS spectra with predicted MS/MS spectra generated from a set of possible proteins in a database using one of the two currently most widely used automated database searching algorithms, Sequest [6] and MASCOT [7]. Finally, a long list of peptides and proteins will be reported in terms of a probability score, as is the case for the MASCOT search engine and a recently modified version of Sequest [13].

---

[a] A portion of this chapter is in preparation for publication as: C. Ji, A. Lo, S. Marcus and L. Li, "Large-Scale Evaluation of the Effect of 2MEGA Labeling for Membrane Proteome Analysis Using Nano-LC ESI QTOF". Dr. S. Marcus grew the cells and fractionated membrane proteins. Mr. A. Lo did desalting using solid phase extraction, strong cation exchange separation, and also helped with the data analysis.

One of the greatest challenges associated with large-scale proteomics using tandem mass spectrometry (MS/MS) and automated database searching is how to determine the reliability of these protein hits; to reduce the number of proteins that are false positives without compromising the number of correct identifications. To reduce false positive identifications, several "rules of thumb" can be used to carry out critical manual evaluation of large-scale proteomic experiment results [14]. However, the required knowledge of the principle issues involved in peptide analysis by MS, and the time-consuming manual evaluation involved has caused the majority of large-scale proteome analysis results published or reported to be without critical manual interpretation. Therefore, any information that can be used to evaluate the reliability of a particular protein identification via an automated method is highly required. Recent reports have begun to raise key issues relating to the confidence of these identifications [15-18]. Additional supporting information from a peptide's LC elution time or its isoelectric point has been developed to reduce further the false positive rates for peptide identifications [19, 20].

Chemical modifications of N-termini and C-termini of peptides have been developed to simplify and direct the fragmentation of peptides and facilitate the interpretation of the obtained MS/MS spectra [21-29]. Divinyl sulfate (DVS) has been used as a post-digestion modifier to enhance the intensity of the signal of the a_1 ion produced in MS/MS and post-source decay [29]. This enhanced signal can be used to fingerprint the N-terminal amino acid of a peptide. This information, which is normally not present in low energy collision induced dissociation (CID) spectra, is advantageous for *de novo* sequencing and should also be used as a filter to reduce false positive identifications. However, DVS may label the N-terminus or certain amino acids (lysine, histidine, cysteine) and produce isomeric products; all these factors complicate CID spectra and hinder the interpretation of peptide sequences. Several studies have reported that dimethyl labeling of amino groups on the N-termini and side chains of lysine residues in the peptide sequences also leads to the enhancement of a_1 ions in the corresponding CID spectra [28, 30-32]. The enhanced a_1 signals in the CID spectra have been demonstrated to provide higher confidence in the identification of proteins performed by either de novo sequencing or database-assisted searching and provide a universal a_1 tag for mapping the N-terminal amino acid through a precursor ion scan with a small set of data [28]. However, the similar masses of a_1 ions derived from N-terminal lysine (157.1705) and arginine (157.1453) residues are indistinguishable when using low-resolution instruments. In addition, the possible multiple labeling of a peptide containing 1~3 lysine residues, which is normally observed in a tryptic digest from a complex protein mixture, will complicate the data analysis. More recently, Reilly reported a novel derivatization strategy, that utilizes both guanidination and amidination, to assist peptide sequencing. A unique characterization of this derivatization is that abundant y_{n-1} and b_1 ions are typically observed in MS/MS spectra [27]. This feature can also be used as a constraint to reduce the false positive identifications. However, it was also reported that the N-terminal amidine groups are susceptible to hydrolysis when the N-terminal residue is serine or threonine. The consequences of this side reaction are that y_{n-1} and b_1 ions were not formed by CID.

In chapter 5, differential 2MEGA labeling of N-termini of peptides with d(0),^{12}C- and d(2),^{13}C-formaldehyde, after blocking the amino groups on the side chains of lysines by guanidination with O-methylisourea, was reported to be a promising strategy for global quantitative and qualitative proteome analysis using auto-offline LC MALDI MS and MS/MS because of the following reasons: (1) the uniform 6.032 Da mass difference between each derivatized peptide pair eliminates the significant overlapping of isotope envelopes even for a peptide pair with m/z around 3 000 Da and simplifies the quantification data analysis process. (2) the reaction itself is simple, fast and complete, and also can be done with commercially available and inexpensive reagents. (3) the presence of universal a1 ions in the MALDI MS/MS spectra and the overlaid fragment ion spectra generated from a pair of differentially labeled peptides can be used to confirm peptide sequences obtained from MS/MS database searching, or to carry out de novo sequencing of peptides based on their MS/MS spectra. In this study, the effect of 2MEGA labeling on the large-scale membrane proteome analysis is further evaluated using nano-LC ESI QTOF mass spectrometry. By comparison with the large-scale membrane proteome analysis of a native digest from the same sample, it is demonstrated that 2MEGA labeling not only increases the number of peptides and proteins but also provides the enhanced a_1 ions or a_1-related ions as a constraint to reduce the number of false positive identifications. The present study will pave the way to further quantitative proteome analysis of membrane proteomes of two different samples prepared under different conditions.

7.2 Experimental

7.2.1 Chemicals and Reagents

d(0),^{12}C-Formaldehyde (37% (w/w) in H_2O), *O*-methylisourea, sodium hydroxide, sodium bicarbonate, sodium cyanoborohydride, bovine trypsin trifluoroacetic acid (TFA) and Leucine enkephalin (Leu-enk) were purchased from Sigma-Aldrich (Oakville, ON, Canada). HPLC grade acetonitrile was purchased from Fisher Scientific

Canada (Edmonton, AB, Canada). Water used in these experiments was obtained from a Milli-Q Plus purification system (Millipore, Bedford, MA). Formic acid and bovine gamma globulin were purchased from Pierce. The other chemicals were from Sigma (St. Louis, MO) and were analytical grade.

7.2.2 Cell Culture and Membrane Preparation

E. coli K-12 (*E. coli*, ATCC 47076) was from the American Type Culture Collection. A single *E. coli* K12 colony was used to inoculate 10 mL of LB broth (BBL, Becton Dickinson). The culture was incubated overnight with shaking at 37 °C. 1.5 mL of this saturated culture was added to 90 mL growth medium in a 500 mL baffled Erlenmeyer flask. Cells were harvested in the mid-log phase by centrifugation at 3 200*g* for 10 min at 4 °C, resuspended, washed in 50 mM MOPS buffer, pH 7.3, and collected by centrifugation at 3 200*g* for 10 min at 4 °C.

A 7 ml aliquot of the E. coli cell suspension was thawed in cold water and the volume was brought to 15 mL with 50 mM MOPS buffer, pH 7.3. Then 1.4 mg of DNaseI was added. The suspension was passed twice through a French press (Amnico Rochester, NY) using rapid fill kit at 14 000 psi. The final volume was about 20 mL after adding more 50 mM MOPS pH 7.3 to rinse the tube. The lysate was centrifuged in a Beckman SX4250 rotor at 4500 rpm (about 2 300 ×g) to pellet unbroken cells for 10 min. The supernatant was collected, and the protein concentration was estimated by performing a BioRad protein assay using bovine gamma globulin as the standard.

The membrane proteins were isolated using a slightly modified carbonate fractionation procedure [33, 34]. About 2 mL lysate (containing approximately 20 mg of cellular proteins) was added to 10 mL ice-cold MOPS buffer. Then, in a 250 mL beaker, 110 mL 0.1 M sodium carbonate pH 11.0 was slowly added. The solution was stirred slowly in an ice bath for 1 hour to extract membranes. The extract was divided equally into two tubes, filled with about 5 mL more 0.1 M sodium carbonate each and centrifuged in a Beckman Type 45Ti rotor for 65 min at 38 400 rpm (115 000 g_{av}). The supernatant was aspirated and the pellet was gently rinsed with 5 mL water. Each pellet was suspended in 2 mL 50 mM MOPS buffer, pH 7.3, and transferred to an 8 mL tube. About 5 mL more buffer was added to each tube to bring the volume to 7 mL. The tubes were centrifuged in a Beckman Type 70.1Ti rotor for 25 min at 40 000 rpm (115 000 g_{av}).

7.2.3 Protein Digestion

To achieve maximum digestion efficiency, two consecutive digestion steps were performed in this study using a combination of organic-assisted [34] and SDS-assisted [35] solubilization and proteolysis. First, proteins in the membrane fraction were resuspended in 50 mM ammonium bicarbonate, pH 8.0, via intermittent vortexing and sonication using a sonicating bath (Bronson model 1510, Danbury CT). The proteins were thermally denatured by incubating the sample in airtight tubes at 90 °C for 20 min, and then cooled down in ice-cold water. The membrane protein concentration was estimated by a BioRad protein assay using bovine gamma globulin as the standard. About 1 mg of protein from the membrane fraction was then diluted with methanol to produce a composition of 60% organic solvent, resulting in a final protein concentration of 1 mg/mL. Tryptic digestion was immediately carried out by adding 20 μg trypsin and incubation at 37 °C for 5 h. Second, after methanol -assisted digestion, un-dissolved sample was pelleted out and re-suspended in 400 μL 0.05% SDS with addition of 15 μg trypsin. The sample was incubated at 37 °C overnight. Methanol in the supernatant was evaporated by SpeedVac and the leftover digest solution was pooled with that from the second digestion. The digestion solution was stored at -80 °C.

7.2.4 2MEGA Labeling

2MEGA labeling of half of the pooled tryptic digest from methanol-assisted and SDS-assisted digestion was carried out as reported previously [36]. In brief, guanidination of lysine residues was performed as described previously [37-40] with some modifications. Trypsin in the 500 μL tryptic digest solution (about 1 μg/μL) was irreversibly inactivated by adding 50 μL 2 M sodium hydroxide. The ε-amino groups of all lysines were blocked by adding 200 μL 2 M *O*-methylisourea in 100 mM $NaHCO_3$, adjusting to pH 11 with 2 M sodium hydroxide and incubating the resulting mixture at 65

°C for 10 min. Then the reaction was stopped and the pH adjusted to 8 by adding 10% TFA. Reductive methylation with d(0), ^{12}C-formaldehyde was also carried out as described previously [28, 30-32] with some modifications. The above guanidinated peptide solution was mixed with 30 μL 2 M sodium cyanoborohydride. The mixture was then vortexed and mixed with d(0),^{12}C-formaldehyde (4% (w/w) in water, 6 μL). The mixture was vortexed and incubated at 37 °C for 1h. If necessary, ammonium bicarbonate (1 M, 6 μL) was added to consume the excess formaldehyde. After labeling, the pH of the solution was adjusted to ~2.5 using 10% TFA. Cautions: sodium cyanoborohydride is a highly toxic compound that releases hydrogen cyanide gas upon exposure to strong acid and formaldehyde is known to have carcinogenic effects, including cancer risk from inhalation exposure. Therefore, the 2MEGA labeling process must be performed in a fume hood.

7.2.5 Desalting Using Solid Phase Extraction Cartridge

The 2MEGA labeled peptide mixture solutions and the unlabeled half of the pooled tryptic digests from the methanol-assisted and SDS-assisted digestions were desalted by SPE using bonded phase octadecyl (C-18) cartridges. Each cartridge was equilibrated with three bed volumes of acetonitrile and washed with three volumes of 0.1% TFA. The peptide mixture was applied to the cartridge and the cartridge was washed with three volumes of aqueous 0.1% TFA. Finally, the peptides were eluted initially with 500 μL of acetonitrile:H_2O:TFA (50:49.9:0.1, v/v/v) and then with 1 mL acetonitrile:H_2O:TFA (75:24.9:0.1, v/v/v). The eluate was concentrated to 300 µL by using a SpeedVac. The peptide mixture was stored at -20 °C.

7.2.6 Cation exchange chromatography

The desalted peptide mixture was separated by strong cation exchange (SCX) chromatography on an Agilent 1100 HPLC system (Palo Alto, CA) using a 2.1 x 150 mm Hydrocell^{TM} SP 1500 column (5 µm, Catalog No.: 24-34 SP, BioChrom Labs, Inc., Terre Haute, IN). The buffer solutions used were 20% v/v acetonitrile in 0.1% TFA (buffer A) and 20% v/v acetonitrile in 0.1% TFA, 1 M NaCl (buffer B). About 400 µL (~500) µg protein digest was loaded onto the SCX column and peptides were eluted with linear gradients of 0-10% B in 2 min, 10-30% B in 10 min and 30-50% B in 2 min at 0.25 mL/min, with collection of 1 min fractions. In total, 8 fractions were collected based the chromatography signal recorded at 214nm. The first two fractions were pooled and the last three fractions were pooled into another fraction because of their low UV absorbance signal. Finally, five fractions were obtained and concentrated to ~10 µL by using a SpeedVac.

7.2.7 Nano-LC ESI QTOF Mass Spectrometric Analysis

The peptides in each SCX fraction were analyzed using a Q-Tof Premier Mass Spectrometer (Waters, Manchester, UK) equipped with a nanoACQUITY^{TM} Ultra Performance LC system (Waters, Milford, MA, USA). In brief, 2 µL peptide solution from each SCX fraction was injected onto a 75µm × 100mm Atlantis™ dC18 column (Waters, Milford, MA, USA). Solvent A consisted of 0.1% formic acid in water; and solvent B consisted of 0.1% formic acid in acetonitrile. Peptides were separated using gradients of 5-30% solvent B in 80 min, 30-90% solvent B in 10 min and 90-5% solvent B in 10 min, and electrosprayed into a Q-Tof Premier Mass Spectrometer, fitted with a nanoLockSpray source, at a flow rate of 250 nL/min. Mass spectra were acquired from *m/z* 300 to 1600 for 1 s followed by 3 data dependent MS/MS scans from 50 to 1900 m/z for 1 s each. The collision energy used to perform MS/MS was varied according to the mass and charge state of the eluting peptide. Leu-enk (Lockmass) was infused at a rate of 250 nL/min and was acquired for 1 s every 2 min throughout the run. The exclusion list was generated based on MASCOT searching results in which peptides with a score above the identity threshold were selected.

7.2.8 Protein Identification from MS/MS Data

Raw search data was lock mass corrected, de-isotoped and converted to peak list files by ProteinLynx Global Server 2.1.5 (Waters). Peptide sequences were identified via automated database searching of peak list files using the MASCOT search program (Matrix Science, London, United Kingdom). Database searching was restricted to *Escherichia coli* in Swiss-Prot database (UniProtKB/Swiss-Prot Release 47.7 of 16-Aug-2005). The following search parameters were selected for all database searching: enzyme, trypsin; missed cleavages, 3; peptide tolerance, ±30 ppm; MS/MS tolerance: 0.2 Da; peptide charge, (1+, 2+ and 3+); variable modification, oxidation (M). In all cases, peak list files were searched twice; in one case with the instrument setting as ESI-QUAD-TOF, the other being constrained to the modified ESI-QUAD-TOF, in which a ions and immonium ions were added as the possible fragment ions. For the database searching of MS/MS data generated from unlabeled pooled tryptic digests from methanol-assisted and SDS-assisted digestion, no additional fixed modifications were selected. However, for the database searching of MS/MS data generated from 2MEGA labeled tryptic peptides from pooled tryptic digests from methanol-assisted and subsequent SDS-assisted digestion, the following modifications were selected as fixed modifications: guanidination (K), Dimethylation-L (N-term) or dimethylation-H (N-term). Where peptides matched more than one database entry due to redundant protein sequence submissions, assignments to the duplicated sequence were removed.

7.2.9 Hydropathy Calculation

All identified proteins were analyzed using the ProtParam tool (available at http://ca.expasy.org/tools/protparam.html), which allows the calculation of the grand average of hydropathy (GRAVY) value for a given protein [41]. The proteins exhibiting positive GRAVY values were recognized as a hydrophobic.

7.3 Results and Discussion

The membrane fraction of *E.coli* cell extract was chosen as the study model for two reasons. First, integral membrane proteins that are inserted into phospholipid bilayers are important biological and pharmacological targets. Second, qualitative, eventually quantitative, large scale proteome analysis of integral membrane proteins remains a challenge. This initial work investigated the effect of 2MEGA labeling on the large-scale proteome analysis of membrane proteins, and will pave the way for later quantitative analysis of membrane proteomes using 2MEGA isotopic labeling. To evaluate the effect of 2MEGA labeling on membrane proteome analysis, large-scale LC-MS/MS datasets for native and 2MEGA labeled tryptic digests from methanol-assisted and subsequent SDS-assisted solubilization and digestion were generated (Figure 7.1). After proteins in the *E. coli* membrane fraction were digested with trypsin using methanol-assisted and subsequent SDS-assisted solubilization (see Experimental section), half of the digest was labeled using the 2MEGA labeling strategy [36]. The native and 2MEGA labeled digests then underwent SPE desalting, SCX separation, RP-nano-LC MS/MS analysis, and database searching.

7.3.1 Fragmentation of Electrosprayed 2MEGA Labeled Peptides

Previous study in Chapter 5 demonstrated that a_1 ions are greatly enhanced after 2MEGA labeling in MALDI MS/MS analysis. In this study, the effect of 2MEGA labeling on the fragmentation of electrosprayed peptides on a large scale is studied further. For this purpose, both unmodified and 2MEGA labeled tryptic peptides, generated from methanol-assisted and subsequent SDS-assisted solubilization of the membrane fraction

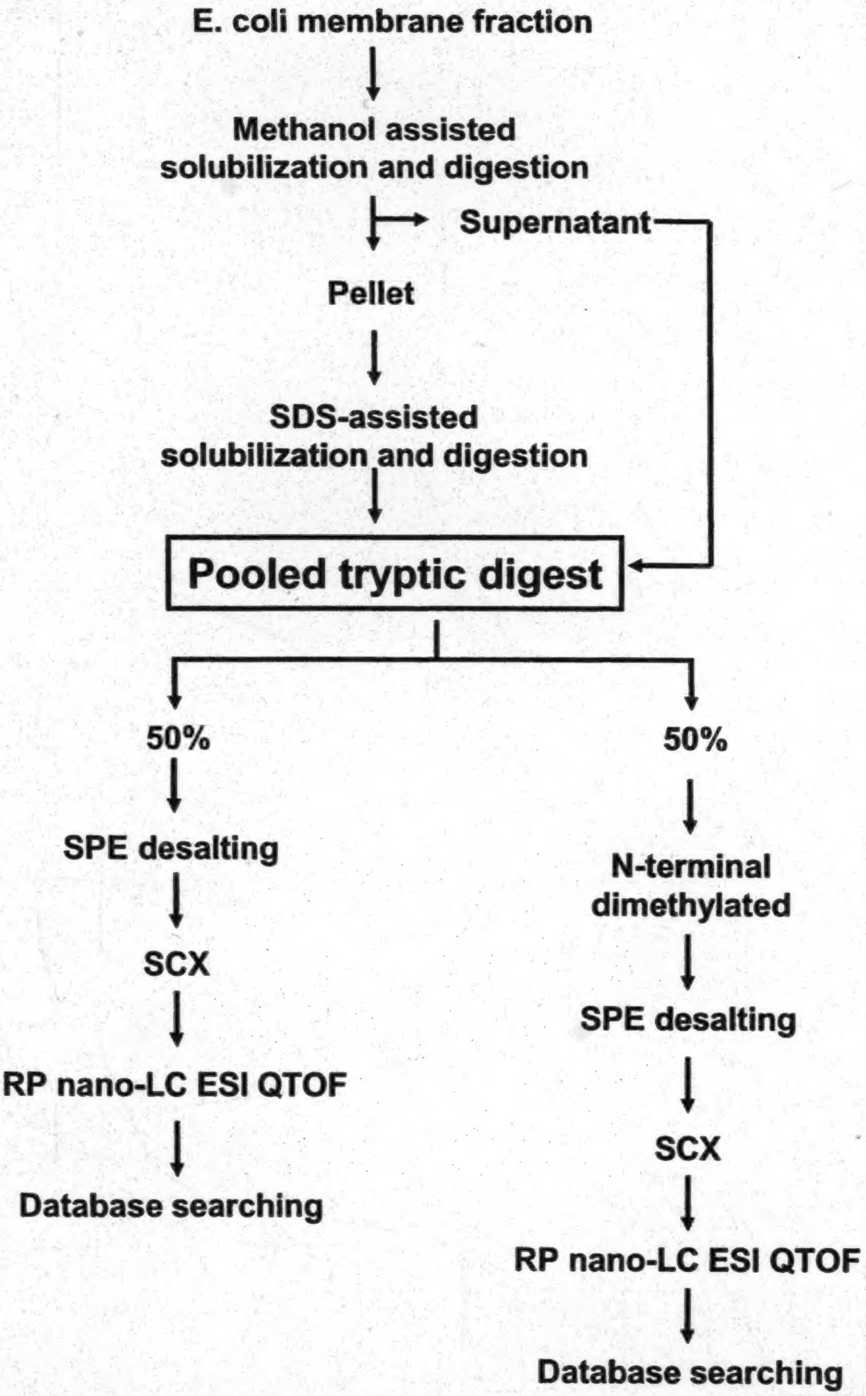
E. coli membrane fraction
Methanol assisted solubilization and digestion
Supernatant
Pellet
SDS-assisted solubilization and digestion
Pooled tryptic digest
50%
50%
SPE desalting
SCX
RP nano-LC ESI QTOF
Database searching
N-terminal dimethylated
SPE desalting
SCX
RP nano-LC ESI QTOF
Database searching

Figure 7.1 Flowchart of the workflow for the investigation of the effect of 2MEGA labeling on the large-scale membrane proteome analysis.

of an *E. coli* cell extract, were analyzed in 2D nLC-MSMS experiments. In total, 1107 unmodified and 1486 2MEGA labeled peptides were identified with MASCOT scores above the identity threshold. Examination of several hundred ESI MSMS spectra of unlabeled and 2MEGA labeled peptide pairs showed that a_1 ion peaks were significantly enhanced in the ESI MS/MS spectra of 2MEGA labeled tryptic peptides. Figure 7.2 displays a pair of representative tandem mass spectra of labeled and unlabeled GYDDEDILK. The a_1 ion peak (30.03) is absent in the spectrum of unlabeled GYDDEDILK (Figure 7.2A), while 2MEGA labeled a_1 ion peak (58.07) is clearly present in the MS/MS spectrum of the 2MEGA labeled GYDDEDILK (Figure 7.2B). In addition, in this case the whole peptide sequence can be easily deduced from the MS/MS spectrum of the 2MEGA labeled peptide (Figure 7.2B). However, one should be mindful that full sequence information can be deduced only if the spectrum is of good quality, such as having reasonably high signal-to-noise ratios for all fragment ions. One advantage of this 2MEGA labeling strategy over the previously reported dimethyl labeling strategy [28] for *de novo* peptide sequencing is that all amino acids are easily distinguished, except L (leucine) and I (isoleucine), whose masses are indistinguishable under low collision energy CID conditions. The previously reported dimethyl labeling strategy cannot distinguish between L/I (leucine/isoleucine) or R/K (arginine/lysine), whose small mass difference (0.025 Da) makes them difficult to resolve, even when using reasonably high mass accuracy instruments, such as QTOF. In most cases, the a_1 ion peak is the strongest peak in the low mass range of the tandem mass spectra of the

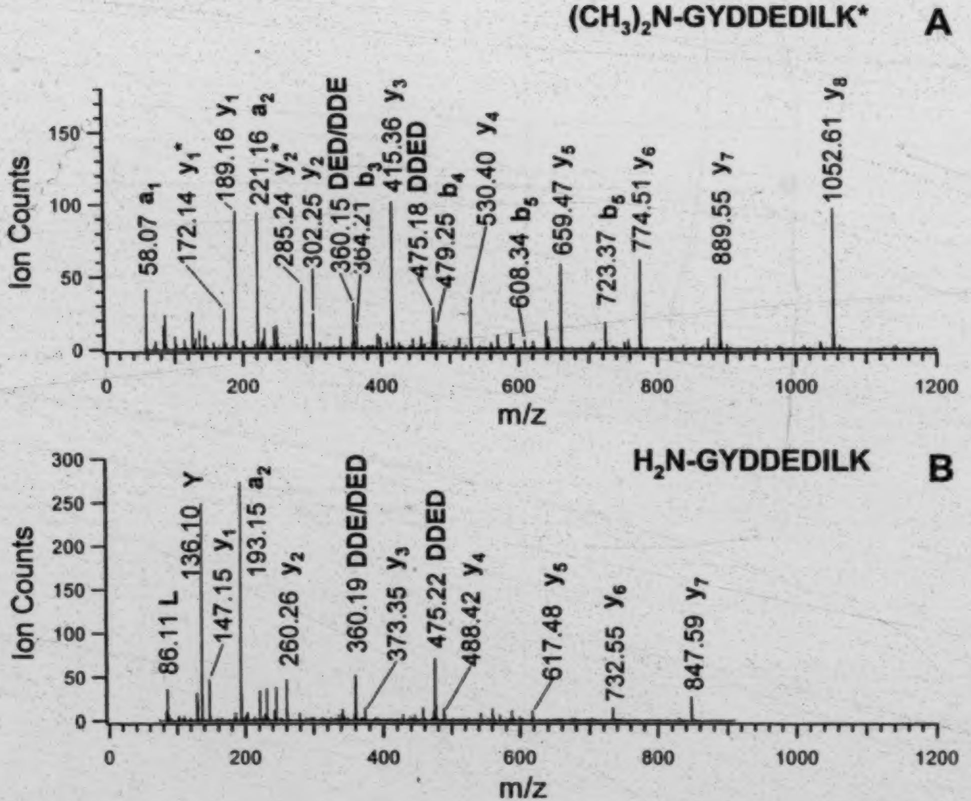

Figure 7.2 ESI MS/MS spectra of (A) 2MEGA labeled and (B) unlabeled GYDDEDILK. * indicates amino group on the side chain of lysine is blocked by guanidination.

2MEGA labeled peptides with different amino acids (AA) as the N-terminal starting AA, except for G, K or R.

Interestingly, instead of observing enhanced a_1 ion peaks, it was found that a_1-45 or a_1-17 peaks are enhanced in ESI MSMS spectra of the 2MEGA labeled tryptic peptides with K or R as the N-terminal starting AA (Figure 7.3), while a_1 ion peaks are often absent or very weak. The observed a_1-17 can be rationalized by the neutral loss of ammonia from the side chain of 2MEGA labeled homoarginine or arginine. The tendency to form a_1-45 in the tandem spectrum of 2MEGA labeled peptides with N-terminal K or R arises from the neutral loss of $(CH_3)_2NH$. This may be due to the fact that the originally formed a_1 positive ions (Figures 7.4A and B) are quickly attacked by one of the lone pairs of electrons on any of three nitrogen atoms on the side chain of 2MEGA labeled homoaginine or arginine to form either a five or seven-membered ring for N-terminal R peptides, or either a six or eight-membered ring for N-terminal K peptides. For 62 identified peptides with N-terminal K and 49 identified peptides with N-terminal R after 2MEGA labeling, at least one of three ions (a_1, a_1-17 and a_1-45) are observed in the corresponding MS/MS spectrum. In most cases, two or all three ions are observed.

Table 7.1 lists the summary of the theoretical masses of a_1 or a_1-related ions from the twenty amino acids, which are commonly observed in the tandem mass spectra of 2MEGA labeled peptides. After checking 1486 tandem spectra of identified peptides from the analysis of an 2MEGA labeled tryptic peptide mixture, using Table 7.1 as the

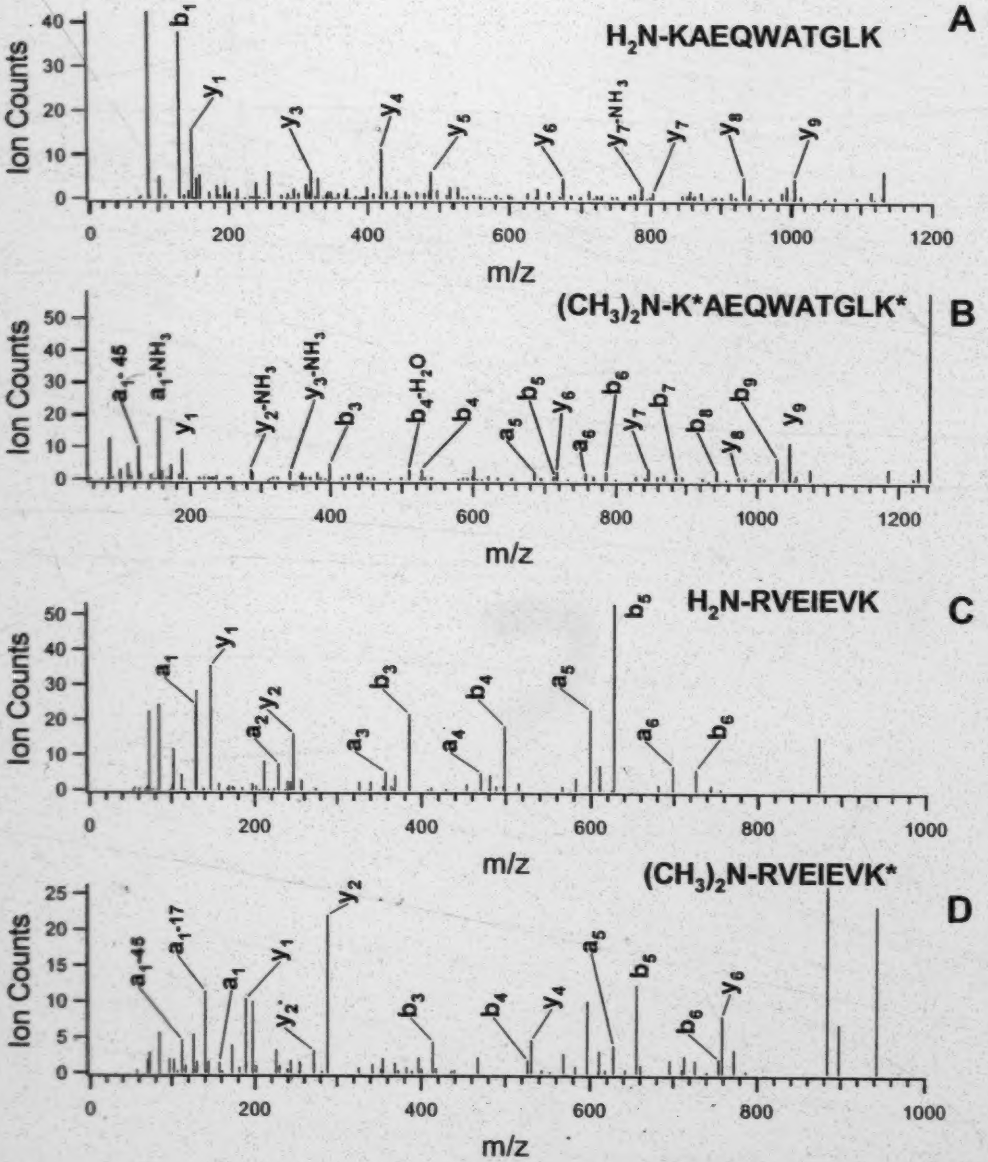

Figure 7.3 ESI MS/MS spectra of various peptides. KAEQWATGLK is shown (A) unlabeled and (B) 2MEGA labeled. RVEIEVK is shown C) unlabeled and D) labeled. * indicates amino group on the side chain of lysine is blocked by guanidination.

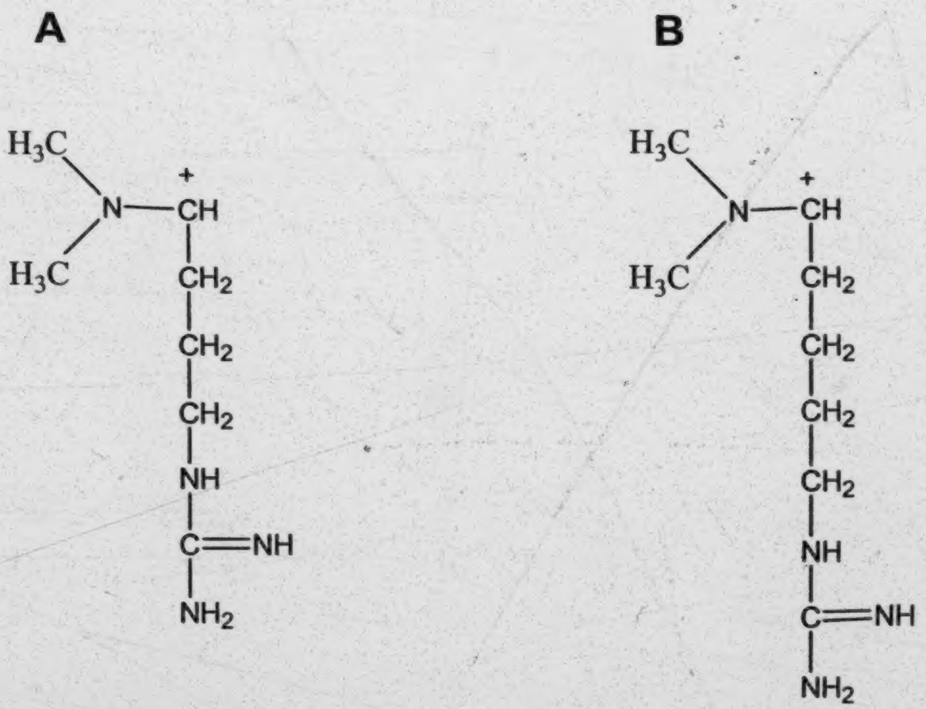
A
B
H3C
H3C
N
+
CH
CH2
NH
C
NH
NH2

Figure 7.4 Representation of the a_1 ion in (A) arginine and (B) homoarginine from guanidinated lysine.

Table 7.1 Theoretical masses of the a_1, a_1-NH_3 and a_1-$HN(CH_3)_2$ ions derived from the twenty amino acid residues after 2MEGA labeling.

N-terminal amino acid residues	Thereoretical a_1 mass	Theoretical a_1-NH_3 mass**	Theoretical a_1-$HN(CH_3)_2$ mass***
	2MEGA labeled		
Alanine (A)	72.081		
Arginine (R)	157.145	140.118	112.087
Asparagine (N)	115.037		
Aspartic acid (D)	116.071		
Cysteine (C)*	161.087		
Glutamic acid (E)	130.087		
Glutamine (Q)	129.103		
Glycine (G)	58.066		
Histidine (H)	138.103		
Isoleucine (I)	114.128		
Leucine (L)	114.128		
Lysine (K)	171.161	154.134	126.103
Methionine (M)	132.085		
Phenylalanine (F)	148.113		
Proline (P)	84.081		
Serine (S)	88.076		
Threonine (T)	102.092		
Tryptophan (W)	187.124		
Tyrosine (Y)	164.108		
Valine (V)	100.113		

* Side chain of cysteine was blocked by iodoacetamide.

** a_1-NH_3 ion peaks only observed for peptides with N-terminal K or R.

*** a_1-$HN(CH_3)_2$ ion peaks only observed for peptides with N-terminal K or R.

reference mass table, a_1 or a_1-related ion peaks are observed in 1461 of them. It was found that a_1 ions (58.07) are not observed in ten tandem spectra of identified peptides with N-terminal G (glycine). This is not surprising because the a_1 ion peak is a relatively weak peak in the low mass range of the tandem mass spectra of the 2MEGA labeled peptides with N-terminal G in which a_1 was observed. Therefore, it is assumed that those are true identifications. Not counting the identified peptides with N-terminal G, 1395 out of 1410 (98.94%) identified peptides with scores above the MASCOT identity threshold have a_1 or a_1-related ions in their MS/MS spectra after 2MEGA labeling. Only 15 of 1410 (1.06%) identified peptides with scores above the MASCOT identity threshold did not have a_1 or a_1-related ions in their MS/MS spectra after 2MEGA labeling, which were discarded as false positive identifications. The low percentage false positive identification rate, calculated on the basis of manually checking the tandem mass spectra using the a_1 or a_1-related ion table (Table 7.1), is consistent with the results reported by Balgley and coworkers [16]. Therefore, a_1 or a_1-related ions can be used as additional information to eliminate false positive identifications for large-scale proteome analysis. Figure 7.5 shows an example of using a_1 or a_1-related ions as a criterion to eliminate the false positive identifications. A good MS/MS spectrum (Figure 7.5A) was searched against the database using MASCOT. Peptide sequence NYQQSYAFVEK was identified as the only significant match with a score well above the identity threshold (Figure 7.4B). Most of the fragment peaks of the identified peptide are well matched with those in the experimental MS/MS spectrum (Figure 7.4A). After manually

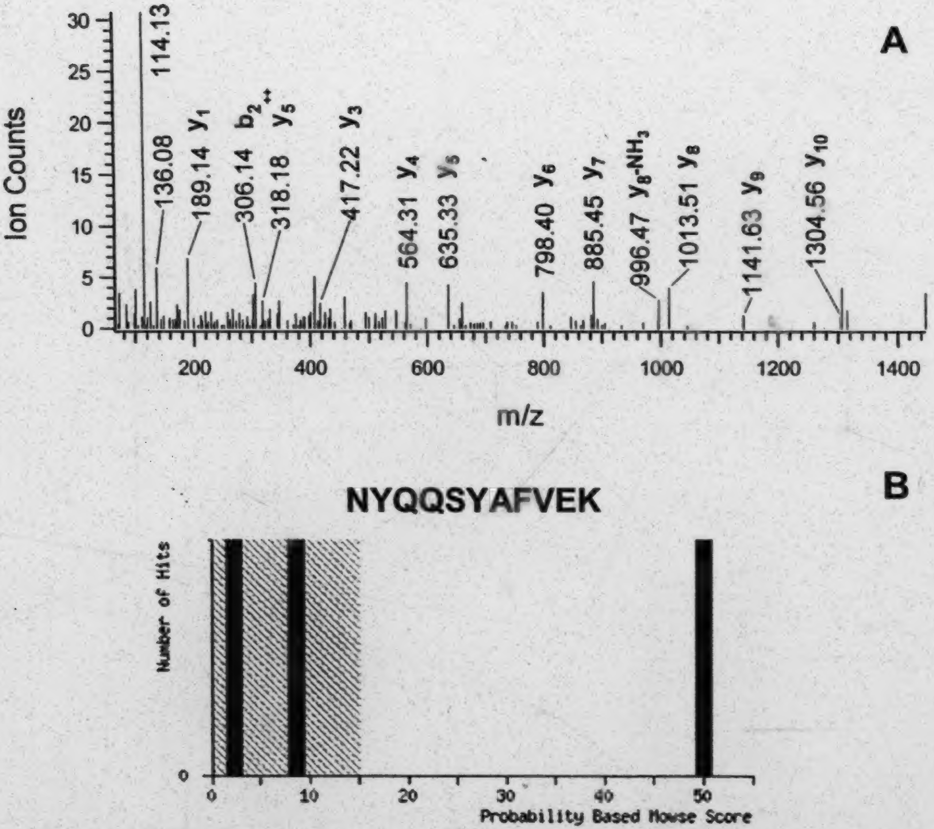

Figure 7.5 An example of detecting false positive identification based on a1 or a1-related ion. (A) ESI MS/MS spectrum and (B) MASCOT search result.

checking the spectrum using the theoretical masses of a_1 or a_1-related ions, the a_1 ion peak with m/z close to 115.037, corresponding to the N-terminal N (aspartic acid), is not observed. However, a strong peak at m/z of 114.13 strongly suggests that the potential true peptide should be one with N-terminal L/I (leucine/isoleucine), which has a theoretical value of 114.128 for the a_1 ion. Therefore, the first match provided by MASCOT searching result could be a false positive identification. However, the second matched peptide (IECPYGPLVEEK) with N-terminal I could be the correct match even though its calculated score is very low.

7.3.2 The Effect of Instrument Settings on Database Searching

Immonium ion peaks and a-series ion peaks are often observed in the MS/MS spectra of labeled and unlabeled tryptic peptides, generated by nLC ESI QTOF MS. In particular, a_1 ion peaks are enhanced after 2MEGA labeling. However, immonium and a-series ions are not contained in the default fragment ions of parent peptides using the default ESI-QUAD-QTOF instrument setting as a MASCOT search parameter. To test the effect of including immonium and a_1 ions as search conditions, those ions were added as possible fragment ions for parent peptides when an ESI QTOF instrument was used to generate CID spectra and generated a end-user defined ESI-QTOF instrument. For the purpose of comparison, both the raw spectral data of labeled and unlabeled tryptic peptides, generated using methanol-assisted and subsequent SDS-assisted solubilization and digestion, were searched against the database using MASCOT twice. The first

Table 7.2 Comparison of database search results from MASCOT using default ESI-QUAD-QTOF without counting immonium ions and a-series ions and custom modified ESI-QUAD-QTOF counting immonium ions and a-series ions.

	Unlabeled		2MEGA labeled	
	ESI-QUAD QTOF	Custom modified ESI-QUAD-QTOF	ESI-QUAD QTOF	Custom modified ESI-QUAD-QTOF
No. of peptides identified	1036	1107	1373(1360)	1486(1471)
No. of proteins identified	397	410	451(447)	502(496)
Average score	55.18	62	55.73	62.68
Score change range	-3~25*		0~25	
Average score increase	6.82		6.95	

* Only two peptides identified with slightly (one case, -1; the other case, -3) reduced scores after counting immonium ions and a-series ions in the unlabeled sample.

** the number inside the bracket indicates the number of peptides or proteins identified after manually checking using a1 or a1-related ions as a constraint.

search was constrained to the default ESI-QUAD-TOF defined in MASCOT software, while a second search was constrained to the end-user defined ESI-QTOF. A summary of the comparison results is listed in Table 7.2. It was generally found that, after counting immonium ions and a-series ions, MASCOT scores for the identified labeled and labeled either remain the same or increase, while the identity threshold remains unchanged. Only two exceptions with minor decreased scores (in those cases, scores decreased by 1 and 3) were observed for unlabeled peptides. The score increase for labeled and unlabeled peptides ranges anywhere from 0 to 25, with an average of 7.11 for labeled peptides, and 6.74 for unlabeled peptides. While counting immonium ions and a-series ions did not greatly increase the average MASCOT scores for the labeled or unlabeled peptides, it did lead to more true identifications. An additional 71 and 113 unique peptides that were initially scored below threshold had their scores increase above threshold for the unlabeled samples and labeled samples, respectively. This represents a percentage increase of 6.85% for the unlabeled samples and 8.23% for the labeled samples. The CID spectra of 113 new identified unique peptides for the labeled samples were manually checked using a_1 or a_1-related ions as the criteria to eliminate the false positive identifications. Of the 113 new peptides, only 2 were discarded as false positive identifications because of the absence of a_1 or a_1-related ions. Therefore, in this study, all the reported identified protein numbers and scores are based on database searching using a custom modified ESI QUAD TOF instrument, unless otherwise noted.

7.3.3 The Effect of 2MEGA Labeling on the Proteome Analysis

Table 7.3 shows a summary of the comparison for all peptides identified by the unlabeled and 2MEGA labeled experiments. There is a significant increase (+32.8%) in the total number of peptides identified in the labeled sample (1471 peptides) versus the unlabeled sample (1107 peptides) and a dramatic increase (84.3%) in the total number of peptides with C-terminal K identified in the labeled sample (645 peptides) versus the unlabeled sample (350 peptides). Both trends can be rationalized by the increased basicity of peptides after guanidination, which selectively converts the amino group on the lysine side chain into a guanidino moiety (identical to the functional group on the arginine side chain). After this conversion, the basicity of homoargine residues is similar to that of arginine, increasing ionization efficiency of peptides with C-terminal K. This explanation also supports the dramatic increase (100%) in the total number of peptides containing K but no R identified in the labeled sample (530 peptides) versus unlabeled sample (265 peptides). These observations are consistent with previous reports that guanidination beneficially increases detection of lysine-terminal peptides in tryptic digest mixtures in MALDI analysis [42]. In addition, there are slight increases (10.4%) in the total number of peptides identified in the labeled sample (815 peptides) versus the unlabeled sample (738 peptides) and (30.9%) in the total number of peptides containing R but no K identified in the labeled sample (466 peptides) versus the unlabeled sample (610 peptides). These observations can not be explained by basicity

Table 7.3 Summary of analysis results of identified peptides generated from membrane fraction of *E.coli* cell extract under labeled and unlabeled conditions.

	Unlabeled		2MEGA labeled	
N-terminal starts with	no. of peptides	Percentage	no. of peptides	Percentage
A	102	9.2	140	9.5
C	0	0.0	0	0.0
D	63	5.7	97	6.6
E	54	4.9	85	5.8
F	56	5.1	73	5.0
G	79	7.1	76	5.2
H	37	3.3	45	3.1
I	72	6.5	115	7.8
K	51	4.6	61	4.1
L	106	9.6	164	11.1
M	50	4.5	57	3.9
N	42	3.8	60	4.1
P	1	0.1	1	0.1
Q	41	3.7	53	3.6
R	52	4.7	48	3.3
S	76	6.9	85	5.8
T	63	5.7	83	5.6
V	71	6.4	115	7.8
W	24	2.2	41	2.8
Y	67	6.1	72	4.9
Total No. of Peptides	1107		1471	
C-terminal end with K	350	31.6	645	43.8
C-terminal end with R	738	66.7	815	55.3
containing K but no R	265	23.9	530	36.0
containing R but no K	466	42.1	610	41.4

alone since the basicity of dimethylated N-terminal amino group is decreased slightly. Two extra methyl groups added to the N-terminal of the peptide could significantly alter the ESI response, which is consistent with that reported by Brancia and coworkers [43]. There is no significant enhancement in the number of peptides with any particular amino acid at the N-terminus, indicating that the 2MEGA labeling can be used to analyze a wide range of samples. So far, it has been demonstrated that 2MEGA labeling not only provides reliable a_1 or a_1-related ions as the additional information for reducing the number of false positive identifications, but also increases the number of peptides identified.

7.3.4 Identification of Membrane Proteins in *E.coli* Membrane Fraction

To ascertain the reproducibility of the overall analysis and the extent to which an increase in the number of unique peptides may be identified per nRPLC-ESI MS/MS analysis, a second nRPLC-MS/MS run of all SCX fractions using the exclusion list resulted in the identification of an average of 20% more unique peptides, and the 3rd analysis of one selected SCX fraction yielded 6% more. Therefore, to save instrument time, each SCX fraction was run twice with the use of the exclusion list for the second run.

Table 7.5 lists all proteins identified from unlabeled and 2MEGA labeled samples with their subcellular locations included. Figure 7.6 shows the number of peptides that were used to identify each protein in 2D LC QTOF analysis of labeled and unlabeled samples. 1471 unique peptides corresponding to 494 unique proteins were unambiguously identified from 2MEGA labeled tryptic peptides of proteins from the membrane fraction of the *E.coli* cell extract, of which, 276 proteins (55.6%) were identified based on two or more peptides (Figure 7.6A). From the unlabeled tryptic peptides, 1107 unique peptides, corresponding to 410 unique proteins, were identified from the membrane fraction of the *E.coli* cell extract, of which, 219 proteins (53.4%) were identified based on two or more peptides (Figure 7.6B). Figure 7.7 illustrates the overlap in proteins identified from unlabeled and 2MEGA labeled samples. Out of 640 proteins identified, 266 proteins are common to both the labeled and unlabeled experiments. These complementary results may indicate that comprehensive proteome analysis results may be achieved by using the combination analysis of unlabeled and 2MEGA samples. In other words, to achieve comprehensive proteome analysis, powerful separation techniques should be developed. In addition, there is a significant increase (21.0%) in the total number of proteins identified in the labeled sample versus unlabeled sample, indicating the promising future of quantitative proteome analysis using differential 2MEGA labeling. The subcellular locations of the identified proteins in the labeled and unlabeled samples illustrated in Swiss-Prot database were investigated further (see Table 7.4). Interestingly, there is a dramatic increase (95.8%) in the total number of integral membrane proteins identified in the 2MEGA labeled sample (141 proteins) versus the unlabeled sample (72 proteins). There are also significant increases in the number (99) and percentage (39.6%) of membrane and membrane-associated proteins

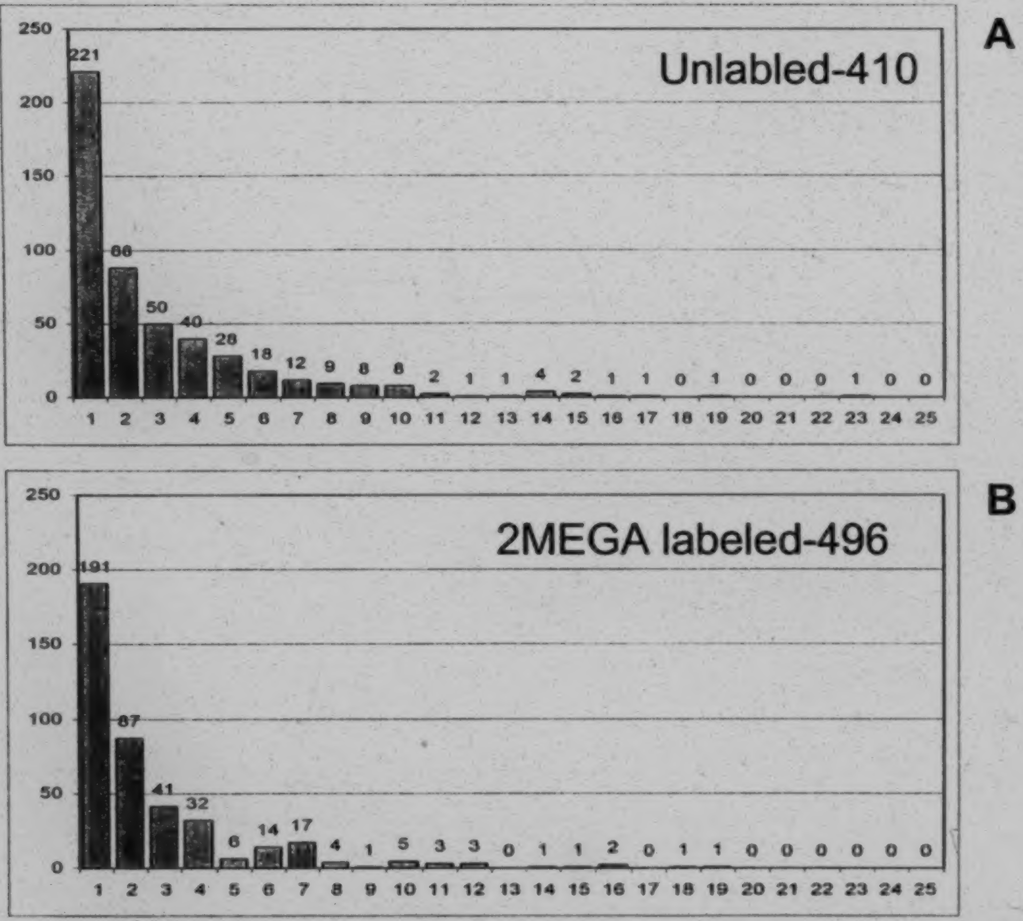

Figure 7.6 Graphs of the total number of proteins versus the number of identifying peptides for (A) unlabeled and (B) 2MEGA labeled samples.

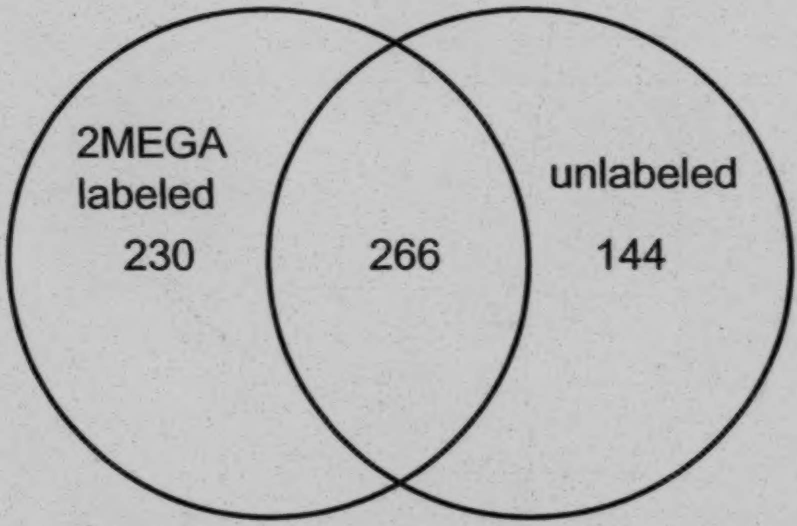

Figure 7.7 Diagram of total proteins identified from unlabeled and 2MEGA labeled tryptic digest generated from membrane fraction of *E. coli* cell extract.

Table 7.4 Subcellular localizations of proteins identified from unlabeled and 2MEGA labeled tryptic digest of proteins from membrane fraction of *E. coli* cell extract.

Subcellular locations	unlabeled	% of total	2MEGA labeled	% of total	common proteins	% of total
Integral membrane (including inner and outer) protein	72	17.4	141	28.1	146.0	22.6
inner membrane (not including those integral membrane protein)	4	1.0	4	0.8	4	0.6
membrane-associated(including inner and outer), bound and anchored	27	6.5	49	9.8	56	8.7
outer membrane (not including those Integral membrane protein)	8	1.9	8	1.6	8	1.2
Attached to the membrane (including inner and outer) by a lipid anchor	26	6.3	28	5.6	31	4.8
membrane	1	0.2	1	0.2	1	0.2
Type I, II and III membrane protein.	5	1.2	11	2.2	12	1.9
Cytoplasmic	56	13.6	39	7.8	75	11.6
Periplasmic.	6	1.5	8	1.6	8	1.2
Restricted to the nucleoid region.	1	0.2	1	0.2	1	0.2
Seems to be associated with the nucleoid (By similarity).	1	0.2	0	0.0	1	0.2
Unknow	206	49.9	211	42.1	303	46.9
total*	413	100.0	501	100.0	646	100.0

* Several proteins have more than one subcellular location in Swiss-Prot database.

identified in the 2MEGA labeled sample (242 and 48.3%) when compared to the unlabeled sample (143 and 34.6%). These results demonstrate that this labeling strategy is an efficient way to identify membrane or membrane-associated proteins. Overall, 258 out of 640 (39.9%) proteins identified in this study are membrane or membrane-associated proteins.

Positive GRAVY values have been considered reliable marker for indicating the hydrophobicity of a protein and a valid indicator of its membrane involvement [41, 44-46]. Figure 7.8 shows the distribution of the number of proteins identified based on their calculated GRAVY. 135 of 496 (27.2%) proteins identified in 2MEGA labeled sample are hydrophobic with positive GRAVY values ranging from +0.001 to +1.271, while 68 of 410 (16.6%) proteins identified in the labeled sample are hydrophobic with positive GRAVY values ranging from +0.001 to +1.121. These results further support the above statement that the 2Mega labeling strategy is an efficient way to identified membrane or membrane-associated proteins. Overall, 154 out of 640 (24.1%) proteins identified in this study are hydrophobic with positive GRAVY values.

7.4 Conclusions

The effect of 2MEGA labeling was evaluated on a large-scale membrane proteome analysis by a shotgun proteomics strategy using the low-parts per million mass accuracy of a QTOF instrument. In this study, it was found that either a_1 ions for

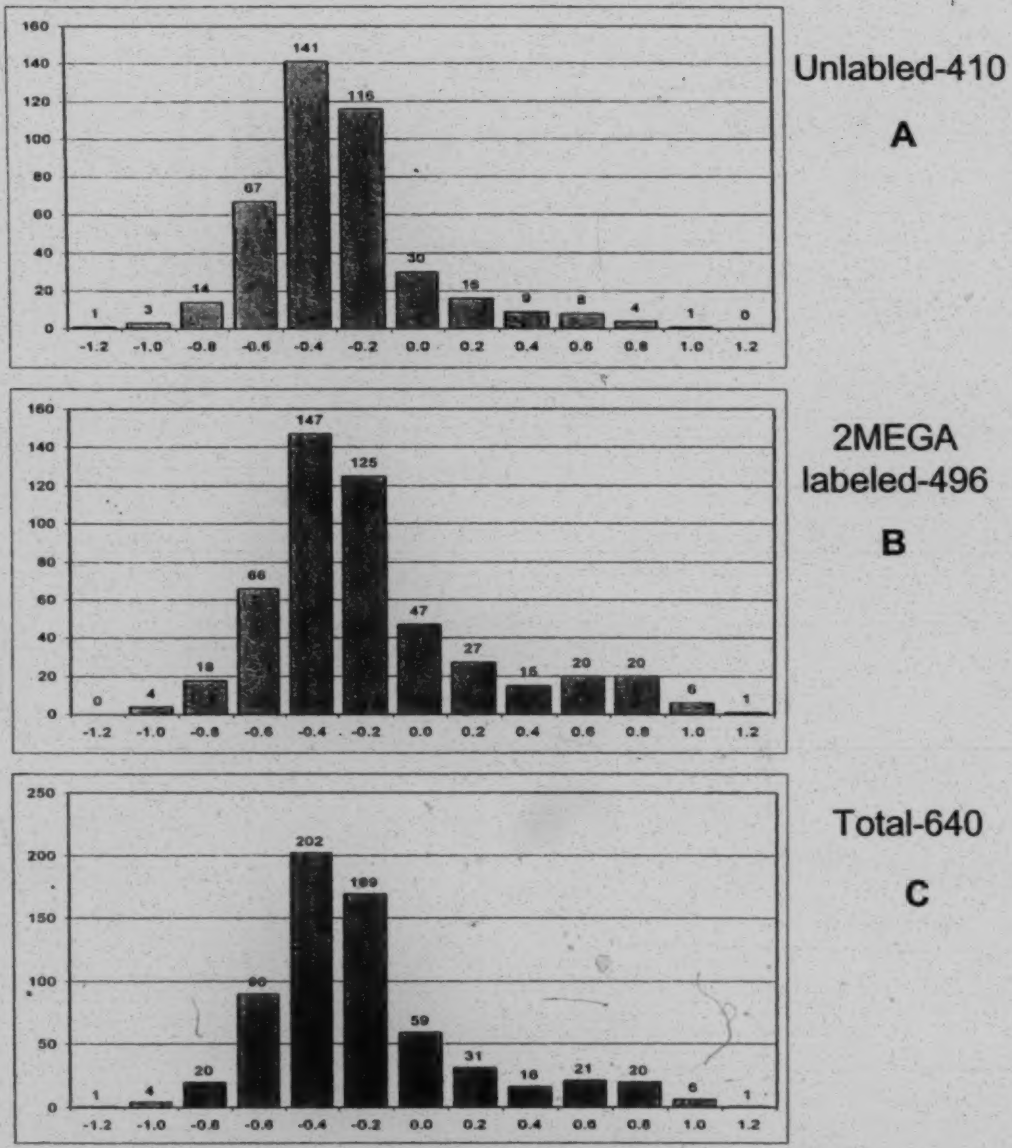

Figure 7.8 Graphs of distributions of calculated GRAVY scores for proteins identified in (A) unlabeled sample and (B) 2MEGA labeled sample, and (C) for total number of unique proteins identified in both unlabeled and 2MEGA labeled samples.

peptides starting with all AA except K and R or a_1-17 or a_1-45 ions for peptides starting with K or R are greatly enhanced when analyzed by ESI MS/MS; these ions are usually difficult to detect in the tandem mass spectra of unlabeled peptides. The 2MEGA labeling strategy alleviated the biased detection of arginine-terminated peptides that is often observed in MALDI and ESI MS experiments. The enhanced $a_1$or a_1-related ions in MS/MS spectra of 2MEGA labeled peptides provide additional information to check spectra and reduce the number of false positive identifications. Although spectra were manually verified in this study, it could be done automatically using a simple program. Based on the data evaluated, about 99% of proteins identified, using MASCOT identity as the threshold, were found to be true identifications. In addition, the immonium ions and a-series ions should be counted as possible fragment ions for parent peptides to reduce the number of false negative identifications when QTOF is used to generated CID spectra. Overall, 640 unique proteins were identified, including 202 membrane proteins and 56 membrane-associated proteins.

7.5 Literature Cited

(1) Hager, J. W. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 512-526.

(2) Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889-896.

(3) Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.;

Vestal, M. L.; Burlingame, A. L. *Anal Chem* **2000**, *72*, 552-558.

(4) Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1047-1057.

(5) Field, H. I.; Fenyo, D.; Beavis, R. C. *Proteomics* **2002**, *2*, 36-47.

(6) Eng, J. K.; McCormack, A. L.; Yates, J. R., 3rd *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976-989.

(7) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.

(8) Shen, Y.; Smith, R. D. *Electrophoresis* **2002**, *23*, 3106-3124.

(9) Washburn, M. P.; Wolters, D.; Yates, J. R., 3rd *Nat. Biotechnol.* **2001**, *19*, 242-247.

(10) Dai, J.; Shieh, C. H.; Sheng, Q.; Zhou, H.; Zeng, R. *Anal. Chem.* **2005**, *77*.

(11) Cargile, B. J.; Talley, D. L.; Stephenson, J. L., Jr. *Electrophoresis* **2004**, *25*, 936-945.

(12) Wolters, D. A.; Washburn, M. P.; Yates, J. R., 3rd *Anal. Chem.* **2001**, *73*, 5683-5690.

(13) MacCoss, M. J.; Wu, C. C.; Yates, J. R., 3rd *Anal Chem* **2002**, *74*, 5593-5599.

(14) Steen, H.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699-711.

(15) Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. *J. Proteome Res.* **2004**, *3*, 1082-1085.

(16) Rudnick, P. A.; Wang, Y.; Evans, E.; Lee, C. S.; Balgley, B. M. *J. Proteome Res.*

2005, *4*, 1353-1360.

(17) Qian, W. J.; Liu, T.; Monroe, M. E.; Strittmatter, E. F.; Jacobs, J. M.; Kangas, L. J.; Petritis, K.; Camp, D. G., 2nd; Smith, R. D. *J. Proteome Res.* **2005**, *4*, 53-62.

(18) Olsen, J. V.; Ong, S. E.; Mann, M. *Mol. Cell. Proteomics* **2004**, *3*, 608-614.

(19) Cargile, B. J.; Bundy, J. L.; Freeman, T. W.; Stephenson, J. L., Jr. *J. Proteome Res.* **2004**, *3*, 112-119.

(20) Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Pasa-Tolic, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E. F.; Shen, Y.; Zhao, R.; Smith, R. D. *Anal. Chem.* **2003**, *75*, 1039-1048.

(21) Cagney, G.; Emili, A. *Nat. Biotechnol.* **2002**, *20*, 163-170.

(22) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 7131-7136.

(23) Keough, T.; Youngquist, R. S.; Lacey, M. P. *Anal. Chem.* **2003**, *75*, 156A-165A.

(24) Lee, Y. ~~H.; Kim~~, M. S.; Choie, W. S.; Min, H. K.; Lee, S. W. *Proteomics* **2004**, *4*, 1684-1694.

(25) Schnolzer, M.; Jedrzejewski, P.; Lehmann, W. D. *Electrophoresis* **1996**, *17*, 945-953.

(26) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun Mass Spectrom* **1997**, *11*, 1015-1024.

(27) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. *Anal. Chem.* **2005**, *77*.

(28) Hsu, J. L.; Huang, S. Y.; Shiea, J. T.; Huang, W. Y.; Chen, S. H. *J. Proteome Res.*

2005, *4*, 101-108.

(29) Boja, E. S.; Sokoloski, E. A.; Fales, H. M. *Anal. Chem.* **2004**, *76*, 3958-3970.

(30) Ji, C.; Li, L.; Gebre, M.; Pasdar, M. *J. Proteome Res.* **2005**, *4*, 1419-1426.

(31) Ji, C.; Li, L. *J. Proteome Res.* **2005**, *4*, 734-742.

(32) Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. *Anal. Chem.* **2003**, *75*, 6843-6852.

(33) Fujiki, Y.; Hubbard, A. L.; Fowler, S.; Lazarow, P. B. *J. Cell Biol.* **1982**, *93*, 97-102.

(34) Blonder, J.; Goshe, M. B.; Moore, R. J.; Pasa-Tolic, L.; Masselon, C. D.; Lipton, M. S.; Smith, R. D. *J. Proteome Res.* **2002**, *1*, 351-360.

(35) Zhang, N.; Li, N.; Li, L. *J. Proteome Res.* **2004**, *3*, 719-727.

(36) Ji, C.; Guo, N.; Li, L. *J. Proteome Res.* **2005**, in press.

(37) Kimmel, J. R. *Methods Enzymol.* **1967**, *11*, 584-589.

(38) Beardsley, R. L.; Reilly, J. P. *Anal. Chem.* **2002**, *74*, 1884-1890.

(39) Zappacosta, F.; Annan, R. S. *Anal. Chem.* **2004**, *76*, 6618-6627.

(40) Brancia, F. L.; Montgomery, H.; Tanaka, K.; Kumashiro, S. *Anal. Chem.* **2004**, *76*, 2748-2755.

(41) Kyte, J.; Doolittle, R. F. *J. Mol. Biol.* **1982**, *157*, 105-132.

(42) Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2070-2073.

(43) Brancia, F. L.; Openshaw, M. E.; Kumashiro, S. *Rapid Commun. Mass Spectrom.*

2002, *16*, 2255-2259.

(44) Nouwens, A. S.; Cordwell, S. J.; Larsen, M. R.; Molloy, M. P.; Gillings, M.; Willcox, M. D.; Walsh, B. J. *Electrophoresis* **2000**, *21*, 3797-3809.

(45) Molloy, M. P.; Herbert, B. R.; Slade, M. B.; Rabilloud, T.; Nouwens, A. S.; Williams, K. L.; Gooley, A. A. *Eur. J. Biochem.* **2000**, *267*, 2871-2881.

(46) Wilkins, M. R.; Gasteiger, E.; Sanchez, J. C.; Bairoch, A.; Hochstrasser, D. F. *Electrophoresis* **1998**, *19*, 1501-1505.

Table 7.5 List of all proteins identified in *E.coli* membrane protein fraction.

#	Protein Name	Access ID	# of Peptides	Subcellular Location
1	Unknown protein from 2D-page spots M62/M63/O3/O9/T35	P39170	26	Outer membrane
2	Outer membrane protein A	P02934	23	Integral membrane protein, outer membrane
3	Organic solvent tolerance protein	P31554	20	Outer membrane
4	Outer membrane protein C	P06996	20	Integral membrane protein, outer membrane
5	Aerobic glycerol-3-phosphate dehydrogenase	P13035	20	Cytoplasmic
6	NADH-quinone oxidoreductase chain C/D	P33599	20	
7	Outer membrane protein F	P02931	17	Integral membrane protein, outer membrane
8	Protease VII	P09169	17	Integral membrane protein, outer membrane
9	Penicillin-binding protein 6	P08506	16	Inner membrane-associated
10	NADH-quinone oxidoreductase chain 3	P33602	16	
11	ATP synthase beta chain	P00824	15	
12	Succinate dehydrogenase flavoprotein subunit	P10444	15	
13	Hypothetical protein ydiJ	P77748	15	
14	Glycerol-3-phosphate acyltransferase	P0A7A7	15	Membrane-bound
15	ATP synthase B chain	P00859	14	
16	PTS system mannitol-specific EIICBA component	P00550	14	Integral membrane protein, inner membrane
17	Vitamin B12 transporter btuB	P06129	14	Integral membrane protein, outer membrane
18	HflC protein	P25661	12	
19	Peptidyl-prolyl cis-trans isomerase D	P77241	12	Type II membrane protein
20	Penicillin-binding protein 5	P04287	12	Inner membrane-associated
21	Elongation factor Tu	P0A6N1	12	Cytoplasmic and membrane-associated
22	Dihydroorotate dehydrogenase	P0A7E1	12	Inner side of the membrane
23	Lipoprotein-34	P21167	11	Attached to the outer membrane by a lipid anchor
24	Peptidoglycan-associated lipoprotein	P07176	11	Attached to the outer membrane by a lipid anchor
25	ATP synthase alpha chain	P00822	11	
26	NADH-quinone oxidoreductase chain F	P31979	11	
27	Elongation factor G	P0A6M8	11	Cytoplasmic
28	DNA-directed RNA polymerase beta' chain	P0A8T7	11	
29	30S ribosomal protein S2	P0A7V0	11	
30	Biodegradative arginine decarboxylase	P28629	10	Cytoplasmic
31	Outer membrane protein tolC	P02930	10	Integral membrane protein, outer membrane

32	Chain length determinant protein	P76372	10	Integral membrane protein, inner membrane
33	Ferrichrome-iron receptor	P06971	10	Integral membrane protein, outer membrane
34	Cytochrome d ubiquinol oxidase subunit I	P11026	10	Integral membrane protein, inner membrane
35	Hypothetical protein ydgA	P77804	10	
36	Cell division protease ftsH	P28691	10	Integral membrane protein, inner membrane
37	50S ribosomal protein L5	P62399	10	
38	Copper-transporting P-type ATPase	Q59385	10	Integral membrane protein
39	Cysteine synthase A	P11096	9	
40	Sulfite reductase	P17846	9	
41	NAD(P) transhydrogenase subunit beta	P07002	9	Integral membrane protein, inner membrane
42	Protein-export membrane protein secD	P19673	9	Integral membrane protein, inner membrane
43	30S ribosomal protein S7	P02359	9	
44	Penicillin-binding protein 1B	P02919	9	Type II membrane protein, inner membrane
45	Acriflavine resistance protein B	P31224	9	Integral membrane protein, inner membrane
46	50S ribosomal protein L2	P60422	9	
47	30S ribosomal protein S3	P0A7V3	9	
48	Outer membrane protein X	P36546	8	Integral membrane protein, outer membrane
49	Possible protease sohB	P24213	8	
50	Succinate dehydrogenase iron-sulfur protein	P07014	8	
51	Ubiquinol oxidase polypeptide I	P18401	8	Integral membrane protein, inner membrane
52	Sensor-like histidine kinase yojN	P39838	8	Integral membrane protein, inner membrane
53	Hypothetical protein yraM	P45464	8	
54	Penicillin-binding protein 1A	P02918	8	Type II membrane protein, inner membrane
55	Small heat shock protein ibpA	P0C054	8	
56	Outer-membrane lipoprotein lolB	P61320	8	Attached to the outer membrane by a lipid anchor
57	Phosphatidylserine decarboxylase proenzyme	P0A8K1	8	
58	Glycerol kinase	P0A6F3	8	
59	PTS system glucose-specific EIICB component	P69786	8	Integral membrane protein, inner membrane
60	D-methionine-binding lipoprotein metQ	P28635	7	Attached to the membrane by a lipid anchor
61	Antigen 43	P39180	7	Outer membrane-associated
62	HemY protein	P09128	7	
63	Polyribonucleotide nucleotidyltransferase	P05055	7	Cytoplasmic
64	Nucleoside-specific channel-forming protein tsx	P22786	7	Integral membrane protein, outer membrane
65	HflK protein	P25662	7	

66	Lipid A export ATP-binding/permease protein msbA	P60752	7	Integral membrane protein, inner membrane
67	Sulfate/thiosulfate import ATP-binding protein cysA	P16676	7	Inner membrane-associated
68	Ribonuclease E	P21513	7	Cytoplasmic
69	Quinoprotein glucose dehydrogenase	P15877	7	Integral membrane protein, inner membrane
70	Hypothetical UPF0169 lipoprotein yfiO	P77146	7	Attached to the membrane by a lipid anchor
71	Transport ATP-binding protein cydD	P29018	7	Integral membrane protein, inner membrane
72	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	P25665	7	
73	Glucans biosynthesis glucosyltransferase H	P62517	7	Integral membrane protein, inner membrane
74	Phosphoenolpyruvate carboxylase	P00864	7	
75	50S ribosomal protein L22	P61175	7	
76	Beta-lactamase TEM	P62593	7	
77	Hypothetical UPF0004 protein yleA	P77645	7	
78	Formate dehydrogenase-O, major subunit	P32176	7	Cytoplasmic
79	Chaperone clpB	P63284	7	Cytoplasmic
80	Putative lipoprotein ybjP	P75818	6	Attached to the membrane by a lipid anchor
81	High-affinity branched-chain amino acid transport ATP-binding protein livF	P22731	6	
82	Ubiquinol oxidase polypeptide II	P18400	6	Integral membrane protein, inner membrane
83	Cell division ATP-binding protein ftsE	P10115	6	
84	Acyl-coenzyme A dehydrogenase	Q47146	6	
85	Probable ABC transporter ATP-binding protein yhbG	P31220	6	
86	Protein asmA	P28249	6	Periplasmic
87	Hypothetical protein yhjG	P37645	6	
88	MltA-interacting protein	P77486	6	Outer membrane
89	Aldehyde-alcohol dehydrogenase	P17547	6	
90	Glutamate synthase	P09831	6	
91	Hypothetical protein ytfM	P39320	6	
92	Cell division protein zipA	P77173	6	Type Ib membrane protein, inner membrane
93	Sulfate adenylyltransferase subunit 1	P23845	6	
94	PTS system mannose-specific EIIAB component	P69797	6	Cytoplasmic
95	30S ribosomal protein S9	P0A7X3	6	
96	Hypothetical lipoprotein ydcL	P64451	6	Attached to the membrane by a lipid anchor
97	Sulfate adenylyltransferase subunit 2	P21156	6	
98	Transcription termination factor rho	P03002	6	
99	DNA-directed RNA polymerase beta chain	P0A8V2	6	

100	AAS bifunctional protein	P31119	6	Integral membrane protein, inner membrane
101	Putative HTH-type transcriptional regulator yfgA	P27434	5	
102	Acriflavine resistance protein A	P31223	5	Attached to the inner membrane by a lipid anchor
103	Formate acetyltransferase 1	P09373	5	Cytoplasmic
104	Hypothetical protein yefG	P37749	5	Cytoplasmic
105	Hypothetical protein yhiI	P37626	5	
106	Pyruvate dehydrogenase E1 component	P06958	5	
107	DamX protein	P11557	5	
108	Sulfite reductase	P38038	5	
109	Signal peptidase I	P00803	5	Integral membrane protein, inner membrane
110	NAD(P) transhydrogenase subunit alpha	P07001	5	Integral membrane protein, inner membrane
111	Lead, cadmium, zinc and mercury transporting ATPase	P37617	5	Integral membrane protein
112	Putative uroporphyrin-III C-methyltransferase	P09127	5	
113	Putative tagatose 6-phosphate kinase gatZ	P37191	5	
114	NADH dehydrogenase	P00393	5	Membrane
115	50S ribosomal protein L16	P02414	5	
116	Sensor protein cpxA	P08336	5	Integral membrane protein, inner membrane
117	Paraquat-inducible protein B	P43671	5	
118	Hypothetical protein yeaG	P77391	5	
119	D-lactate dehydrogenase	P06149	5	Membrane bound
120	Major outer membrane lipoprotein	P69776	5	Outer membrane
121	Adenylosuccinate synthetase	P0A7D4	5	Cytoplasmic
122	Asparaginyl-tRNA synthetase	P0A8M0	5	Cytoplasmic
123	Maltoporin	P02943	5	Integral membrane protein, outer membrane
124	Lipoprotein-28	P04846	5	Attached to the inner membrane by a lipid anchor
125	D-methionine transport ATP-binding protein metN	P30750	5	Inner membrane-associated
126	30S ribosomal protein S4	P0A7V8	5	
127	Phosphoglycerol transferase I	P39401	5	Integral membrane protein, inner membrane
128	Transport ATP-binding protein cydC	P23886	5	Integral membrane protein, inner membrane
129	High-affinity branched-chain amino acid transport ATP-binding protein livG	P22730	5	
130	1-acyl-sn-glycerol-3-phosphate acyltransferase	P26647	5	Inner membrane-associated
131	Lipid A biosynthesis	P24205	5	Inner membrane-anchored
132	PTS system N-acetylglucosamine-specific EIICBA component	P09323	5	Integral membrane protein, inner membrane
133	DNA protection during starvation protein	P27430	4	

134	Outer membrane protein slp	P37194	4	Attached to the outer membrane by a lipid anchor
135	Phosphoenolpyruvate-protein phosphotransferase ptsP	P37177	4	Cytoplasmic
136	Hypothetical protein yibN	P37688	4	
137	Colicin I receptor	P17315	4	Outer membrane
138	Protease IV	P08395	4	Integral membrane protein, inner membrane
139	Hypothetical protein yagU	P77262	4	Integral membrane protein
140	Phospholipase A1	P00631	4	Outer membrane
141	HTH-type transcriptional regulator malT	P06993	4	
142	Probable N-acetylmuramoyl-L-alanine amidase ybjR	P75820	4	
143	Glutamine transport ATP-binding protein glnQ	P10346	4	Inner membrane-associated
144	ADP-L-glycero-D-manno-heptose-6-epimerase	P67910	4	
145	Hypothetical protein yhcB	P39436	4	
146	Hypothetical protein yfgL	P77774	4	
147	Hypothetical lipoprotein yajG	P36671	4	Attached to the membrane by a lipid anchor
148	Hypothetical UPF0070 protein yfgM	P76576	4	
149	Peptidoglycan synthetase ftsI	P04286	4	Inner membrane Periplasm
150	Lipoprotein nlpI	P39833	4	Attached to the membrane by a lipid anchor
151	Hypothetical ABC transporter ATP-binding protein yhiH	P37624	4	Integral membrane protein, inner membrane
152	Sensor protein phoQ	P23837	4	Integral membrane protein, inner membrane
153	ATP synthase gamma chain	P00837	4	
154	Hypothetical ABC transporter ATP-binding protein yliA	P75796	4	
155	Sensor protein baeS	P30847	4	Integral membrane protein, inner membrane
156	Glycerol-3-phosphate transporter	P08194	4	Integral membrane protein, inner membrane
157	Phosphate import ATP-binding protein pstB	P07655	4	Inner membrane-associated
158	Sensor kinase protein rcsC	P14376	4	Integral membrane protein, inner membrane
159	Hypothetical protein ytfB	P39310	4	
160	Chromosome partition protein mukB	P22523	4	Restricted to the nucleoid region
161	Galactitol-specific phosphotransferase enzyme IIA component	P69828	4	Cytoplasmic
162	Long-chain fatty acid transport protein	P10384	4	Integral membrane protein, outer membrane
163	Hypothetical protein yqjD	P64581	4	
164	Galactitol permease IIC component	P69831	4	Integral membrane protein, inner membrane
165	30S ribosomal protein S18	P0A7T7	4	
166	ATP-dependent RNA helicase rhlB	P0A8J8	4	
167	Inner membrane protein oxaA	P25714	4	Integral membrane protein, inner membrane

168	Maltose/maltodextrin import ATP-binding protein malK	P68187	4	Inner membrane-associated
169	RecA protein	P0A7G6	4	Cytoplasmic
170	Cation efflux system protein cusA	P38054	4	Integral membrane protein, inner membrane
171	Hypothetical UPF0004 protein yliG	P75802	4	
172	Cyclopropane-fatty-acyl-phospholipid synthase	P30010	4	Cytoplasmic
173	Stringent starvation protein A	P05838	4	
174	Lipoyl synthase	P60716	4	Cytoplasmic
175	Hydrogenase-1 small chain	P69739	4	Membrane-bound; periplasmic side
176	Formamidopyrimidine-DNA glycosylase	P05523	4	
177	Cold-shock DEAD-box protein A	P23304	4	Cytoplasmic
178	Translation initiation factor IF-2	P0A705	4	Cytoplasmic
179	UDP-N-acetylglucosamine--N-acetylmuramyl-	P17443	4	Inner membrane-associated
180	Isocitrate dehydrogenase	P08200	4	
181	Hypothetical lipoprotein yfhM	P76578	4	Attached to the membrane by a lipid anchor
182	Bactoprenol glucosyl transferase homolog from prophage CPS-53	P77293	4	
183	Hypothetical protein yefI	P37751	4	Inner membrane-associated
184	Na(+)/H(+) antiporter 2	P27377	4	Integral membrane protein, inner membrane
185	Aspartate-semialdehyde dehydrogenase	P00353	3	
186	Hypothetical symporter ydjN	P77529	3	Integral membrane protein
187	Phage shock protein B	P23854	3	Inner membrane
188	50S ribosomal protein L15	P02413	3	
189	Outer membrane lipoprotein blc	P39281	3	Attached to the outer membrane by a lipid anchor
190	Cell division protein ftsA	P06137	3	
191	ATP synthase delta chain	P00831	3	
192	Hypothetical protein ygiM	P39202	3	
193	Tagatose-1,6-bisphosphate aldolase gatY	P37192	3	
194	Hypothetical protein yjgP	P39340	3	Integral membrane protein
195	Signal recognition particle protein	P07019	3	
196	Penicillin-binding protein 2	P08150	3	Inner membrane
197	Iron(III) dicitrate transport protein fecA	P13036	3	Outer membrane
198	Carbamoyl-phosphate synthase large chain	P00968	3	
199	Ribose transport ATP-binding protein rbsA	P04983	3	Inner membrane-associated
200	Protein-export membrane protein secF	P19674	3	Integral membrane protein, inner membrane
201	NADH-quinone oxidoreductase chain I	P33604	3	

202	NADH-quinone oxidoreductase chain B	P33598	3	
203	Electron transport complex protein rnfG	P77285	3	Inner membrane
204	Methionine synthase	P13009	3	
205	PTS system fructose-specific EIIBC component	P20966	3	Integral membrane protein, inner membrane
206	UPF0141 protein yijP	P32678	3	Integral membrane protein
207	Multidrug resistance protein A	P27303	3	Inner membrane-bound
208	ElaB protein	P52084	3	
209	30S ribosomal protein S1	P02349	3	
210	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	P06959	3	
211	Carbon starvation protein A	P15078	3	Integral membrane protein, inner membrane
212	Hypothetical protein yiaF	P37667	3	
213	Phosphate transport system protein phoU	P07656	3	Cytoplasmic
214	Hypothetical protein yebT	P76272	3	
215	Preprotein translocase secA subunit	P10408	3	Cytoplasmic side of plasma membrane
216	3-deoxy-D-manno-octulosonic-acid transferase	P23282	3	Inner membrane-anchored
217	30S ribosomal protein S10	P0A7R5	3	
218	50S ribosomal protein L4	P60723	3	
219	Large-conductance mechanosensitive channel	P0A742	3	Integral membrane protein, inner membrane
220	Probable UDP-N-acetyl-D-mannosaminuronic acid transferase	P27836	3	
221	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase	P77690	3	
222	PTS-dependent dihydroxyacetone kinase, phosphotransferase subunit dhaM	P37349	3	
223	Rare lipoprotein B	P10101	3	Attached to the membrane by a lipid anchor
224	Membrane-bound lytic murein transglycosylase A	P46885	3	Attached to the outer membrane by a lipid anchor
225	Exodeoxyribonuclease X	P76281	3	
226	Hydrogenase-2 large chain	P37181	3	Membrane-bound
227	Hypothetical protein ycgF	P75990	3	
228	Hypothetical UPF0042 protein yhbJ	P0A894	3	
229	Adenosine deaminase	P22333	3	
230	GTP-binding protein typA/BipA	P32132	3	
231	Aspartate--ammonia ligase	P00963	3	Cytoplasmic
232	CTP synthase	P0A7E5	3	
233	Threonyl-tRNA synthetase	P0A8M3	3	Cytoplasmic
234	Potassium efflux system kefA	P77338	3	Integral membrane protein

235	CDP-diacylglycerol--serine O-phosphatidyltransferase	P23830	3	Cytoplasmic; possible interaction with the inner membrane
236	Dipeptide transport system permease protein dppB	P37316	3	Integral membrane protein, inner membrane
237	Hypothetical protein yhjU	P37659	3	
238	Phosphate regulon sensor protein phoR	P08400	3	Integral membrane protein, inner membrane
239	Methyl-accepting chemotaxis protein I	P02942	3	Integral membrane protein, inner membrane
240	Cystine-binding periplasmic protein	P39174	3	Periplasmic
241	NADH-quinone oxidoreductase chain H	P33603	3	Integral membrane protein
242	Cardiolipin synthetase	P0A6H9	3	Membrane-bound
243	Cation efflux system protein cusB	P77239	3	
244	Cobalamin synthase	P36561	3	
245	GTP-binding protein lepA	P60785	3	
246	Apolipoprotein N-acyltransferase	P23930	3	Integral membrane protein, inner membrane
247	Macrolide-specific efflux protein macA	P75830	3	Inner membrane-associated
248	50S ribosomal protein L20	P0A7L3	3	
249	Rod shape-determining protein mreB	P13519	2	
250	Acyl-CoA thioester hydrolase ybgC	P46130	2	Attached to the outer membrane by a lipid anchor
251	50S ribosomal protein L17	P02416	2	
252	Hypothetical protein yfdC	P37327	2	
253	Hypothetical protein yicH	P31433	2	
254	Hypothetical amino-acid ABC transporter ATP-binding protein yecC	P37774	2	Inner membrane-associated
255	Hypothetical lipoprotein yedD	P31063	2	Attached to the membrane by a lipid anchor
256	Low-affinity inorganic phosphate transporter 1	P37308	2	Integral membrane protein, inner membrane
257	Hypothetical protein yjcD	P32702	2	
258	Hypothetical ABC transporter ATP-binding protein yadG	P36879	2	
259	Hypothetical protein ygaM	Q47413	2	
260	Hypothetical protein yciW	P76035	2	
261	Sensor protein kdpD	P21865	2	Integral membrane protein, inner membrane
262	Copper homeostasis protein cutF	P40710	2	Attached to the outer membrane by a lipid anchor
263	TolQ protein	P05828	2	Integral membrane protein, inner membrane
264	Cell division protein ftsZ	P06138	2	Inner surface of the cytoplasmic membrane
265	Hypothetical protein yraR	P45469	2	
266	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	P30867	2	
267	Aconitate hydratase 2	P36683	2	

268	Hypothetical protein yciS	P77614	2	Integral membrane protein
269	VacJ lipoprotein	P76506	2	Attached to the outer membrane by a lipid anchor
270	Hypothetical UPF0092 protein yajC	P19677	2	
271	Sulfate-binding protein	P06997	2	Periplasmic
272	Hypothetical UPF0003 protein yjeP	P39285	2	Integral membrane protein
273	Sigma-E factor negative regulatory protein	P38106	2	
274	Oligopeptide transport system permease protein oppB	P31132	2	Integral membrane protein, inner membrane
275	Hypothetical protein ydbH	P52645	2	
276	Inositol-1-monophosphatase	P22783	2	
277	30S ribosomal protein S14	P02370	2	
278	Anaerobic C4-dicarboxylate transporter dcuA	P04539	2	Integral membrane protein, inner membrane
279	Hypothetical UPF0003 protein yggB	P11666	2	Integral membrane protein
280	Unknown protein from 2D-page	P39169	2	
281	30S ribosomal protein S16	P0A7T3	2	
282	Probable protease htpX	P23894	2	Integral membrane protein
283	Putative membrane protein igaA homolog	P45800	2	Integral membrane protein, inner membrane
284	ATP synthase epsilon chain	P0A6E6	2	
285	S-adenosylmethionine synthetase	P0A817	2	Cytoplasmic
286	Hypothetical lipoprotein yddW	P64426	2	Attached to the membrane by a lipid anchor
287	Putative transport protein ybjL	P60869	2	Integral membrane protein
288	Chaperone protein dnaK	P0A6Y8	2	
289	Hypothetical protein yraP	P64596	2	Periplasmic
290	Hypothetical UPF0061 protein ydiU	P77649	2	
291	Selenide, water dikinase	P16456	2	
292	Threonine synthase	P00934	2	
293	Carbonic anhydrase 2	P61517	2	
294	Phosphatase yqaB	P77475	2	
295	Single-stranded-DNA-specific exonuclease recJ	P21893	2	
296	Dihydrolipoyl dehydrogenase	P00391	2	Cytoplasmic
297	Maltodextrin phosphorylase	P00490	2	
298	Hypothetical UPF0063 protein yfgB	P36979	2	
299	Hypothetical chaperone protein yegD	P36928	2	
300	Phosphatase yieH	P31467	2	
301	Transcription elongation protein nusA	P03003	2	

302	Phosphoenolpyruvate synthase	P23538	2	
303	DNA gyrase subunit B	P06982	2	
304	Chaperone protein dnaJ	P08622	2	Cytoplasmic
305	Universal stress protein G	P39177	2	
306	tRNA delta	P16384	2	
307	Uridine kinase	P0A8F4	2	Cytoplasmic
308	Hypothetical protein ygcF	P64554	2	
309	Exoribonuclease II	P30850	2	
310	Lysyl-tRNA synthetase, heat inducible	P0A8N5	2	Cytoplasmic
311	3-isopropylmalate dehydratase large subunit	P0A6A6	2	
312	Nucleoid-associated protein ndpA	P33920	2	Associated with the nucleoid
313	Hypothetical protein yjgQ	P39341	2	Integral membrane protein
314	Dipeptide transport ATP-binding protein dppF	P37313	2	Inner membrane-associated
315	Hypothetical protein ygaP	P55734	2	Integral membrane protein
316	Hypothetical protein yciM	P45576	2	
317	Ribonuclease R	P21499	2	
318	Protein yhjK	P37649	2	
319	Hypothetical protein ybbK	P77367	2	
320	Ferrous iron transport protein B	P33650	2	Integral membrane protein, inner membrane
321	DedA protein	P09548	2	Integral membrane protein
322	Glyceraldehyde-3-phosphate dehydrogenase A	P06977	2	Cytoplasmic
323	Dipeptide transport ATP-binding protein dppD	P37314	2	Inner membrane-associated
324	Cation efflux system protein cusC	P77211	2	Attached to the outer membrane by a lipid anchor
325	Preprotein translocase secY subunit	P03844	2	Integral membrane protein, inner membrane
326	Hypothetical protein yedQ	P76330	2	Integral membrane protein
327	Protein yrbC	P45390	2	
328	Hypothetical lipoprotein yeaY	P76255	2	Attached to the membrane by a lipid anchor
329	Tail-specific protease	P23865	2	Periplasmic side of the cytoplasmic membrane
330	Hypothetical tRNA/rRNA methyltransferase yfiF	P33635	2	
331	Histidine transport ATP-binding protein hisP	P07109	2	Inner membrane-associated
332	Peroxidase/catalase HPI	P13029	2	
333	Lipid A biosynthesis lauroyl acyltransferase	P24187	2	Inner membrane-anchored
334	Sensor protein barA	P26607	2	Integral membrane protein, inner membrane
335	Cell division protein ftsL	P22187	2	Type II membrane protein, inner membrane

336	Hypothetical protein yceG	P28306	2	
337	Glutamine transport system permease protein glnP	P10345	2	Integral membrane protein, inner membrane
338	Cell division protein ftsX	P10122	2	Integral membrane protein
339	Lipopolysaccharide 1,2-glucosyltransferase	P27129	2	
340	Protein ycaC	P21367	2	
341	Rod shape-determining protein rodA	P15035	2	Integral membrane protein, inner membrane
342	5,10-methylenetetrahydrofolate reductase	P00394	2	
343	Hypothetical protein yqjK	Q47710	2	
344	NADH-quinone oxidoreductase chain M	P31978	2	Integral membrane protein
345	Sodium/proline symporter	P07117	2	Integral membrane protein, inner membrane
346	Trk system potassium uptake protein trkA	P23868	2	Inner membrane
347	Hypothetical protein ybdG	P39455	2	Integral membrane protein
348	Hypothetical protein ygiC	P24196	2	
349	Taurine import ATP-binding protein tauB	Q8X5I6	2	Inner membrane-associated
350	Alkanesulfonate monooxygenase	P80645	2	
351	Hypothetical ABC transporter ATP-binding protein yrbF	P63386	2	
352	30S ribosomal protein S20	P0A7U7	2	
353	Malate dehydrogenase	P61889	2	
354	Enolase	P0A6P9	2	Cytoplasmic
355	Mannose permease IID component	P69805	2	Integral membrane protein, inner membrane
356	Multidrug resistance protein mdtF	P37637	2	Integral membrane protein, inner membrane
357	Low affinity potassium transport system protein kup	P63183	2	Integral membrane protein, inner membrane
358	Sec-independent protein translocase protein tatA	P69428	2	Inner membrane-bound
359	Osmotically inducible lipoprotein E	P23933	1	Attached to the membrane by a lipid anchor
360	Magnesium and cobalt transport protein corA	P27841	1	Integral membrane protein
361	ATP synthase a chain	P00855	1	Integral membrane protein, inner membrane
362	Glutamine synthetase	P06711	1	Cytoplasmic
363	Hypothetical protein yjeI	P39278	1	
364	Lipopolysaccharide biosynthesis protein wzzE	P25905	1	Integral membrane protein, inner membrane
365	Hypothetical protein yfhG	P37328	1	
366	Thiol:disulfide interchange protein dsbE	P33926	1	Periplasmic; anchored in the inner membrane
367	Bor lipoprotein homolog from lambdoid prophage DLP12	P77330	1	Attached to the membrane by a lipid anchor
368	Lipoprotein spr	P77685	1	Attached to the membrane by a lipid anchor
369	HTH-type transcriptional regulator cbl	Q47083	1	

370	Ferrienterobactin receptor	P05825	1	Integral membrane protein, outer membrane
371	Probable formate transporter 1	P21501	1	Integral membrane protein, inner membrane
372	Outer membrane protein W	P21364	1	Outer membrane
373	Lipoprotein nlpD	P33648	1	Attached to the inner membrane by a lipid anchor
374	Phosphate-binding periplasmic protein	P06128	1	Periplasmic
375	FMN reductase	P80644	1	
376	Aminopeptidase ypdF	P76524	1	
377	Isocitrate lyase	P05313	1	Cytoplasmic
378	Hypothetical UPF0053 protein yfjD	P37908	1	Integral membrane protein
379	Hypothetical protein ycfL	P75946	1	
380	Ubiquinone/menaquinone biosynthesis methyltransferase ubiE	P0A887	1	
381	ATP-dependent Clp protease proteolytic subunit	P0A6G7	1	Cytoplasmic
382	Ribose-phosphate pyrophosphokinase	P0A717	1	Cytoplasmic
383	Protein rcsF	P69411	1	Outer membrane
384	6-phosphofructokinase isozyme I	P0A796	1	Cytoplasmic
385	Aerobic respiration control sensor protein arcB	P22763	1	Integral membrane protein, inner membrane
386	Hypothetical protein yrbD	P64604	1	
387	Chaperone protein htpG	P0A6Z3	1	Cytoplasmic
388	50S ribosomal protein L3	P60438	1	
389	CDP-diacylglycerol pyrophosphatase	P06282	1	Inner membrane-associated
390	UDP-glucose 6-dehydrogenase	P76373	1	
391	Hypothetical protein yegR	P76406	1	
392	ATP-dependent RNA helicase srmB	P21507	1	
393	Bifunctional polymyxin resistance arnA protein	P77398	1	
394	Outer membrane pore protein E	P02932	1	Integral membrane protein, outer membrane
395	Pyruvate formate-lyase 1 activating enzyme	P09374	1	Cytoplasmic
396	Fructose repressor	P21168	1	
397	DNA polymerase III alpha subunit	P10443	1	Cytoplasmic
398	Membrane-bound lytic murein transglycosylase C	P52066	1	Attached to the membrane by a lipid anchor
399	Asparagine synthetase B	P22106	1	
400	Hypothetical protein yjiA	P24203	1	
401	Probable dimethyl sulfoxide reductase chain ynfF	P77783	1	Cytoplasmic face of the membrane
402	SanA protein	P33017	1	Integral membrane protein, inner membrane
403	Aromatic-amino-acid aminotransferase	P04693	1	Cytoplasmic

404	Aconitate hydratase 1	P25516	1	
405	Sensor kinase cusS	P77485	1	Integral membrane protein, inner membrane
406	Galactitol-specific phosphotransferase enzyme IIB component	P37188	1	Cytoplasmic
407	Hypothetical protein yjbB	P32683	1	Integral membrane protein
408	Protein inaA	P27294	1	
409	Protein ycgK	P76002	1	
410	Phosphomethylpyrimidine kinase	P76422	1	
411	Hypothetical protein ycbL	P75849	1	
412	Trehalose-6-phosphate hydrolase	P28904	1	Cytoplasmic
413	NAD(P)H-flavin reductase	P23486	1	
414	Hypothetical protein yfaZ	P76471	1	
415	Hypothetical protein ybgK	P75745	1	
416	GTP-dependent nucleic acid-binding protein engD	P31216	1	
417	Hypothetical protein ygdH	P37350	1	
418	Hypothetical protein yigA	P23305	1	
419	Acetolactate synthase isozyme II large subunit	P00892	1	
420	UTP--glucose-1-phosphate uridylyltransferase	P25520	1	
421	Probable pyruvate-flavodoxin oxidoreductase	P52647	1	
422	Hypothetical protein yajB	P21515	1	
423	FhuE receptor	P16869	1	Outer membrane
424	Putative deoxyribonuclease ycfH	P37346	1	
425	Anaerobic dimethyl sulfoxide reductase chain A	P18775	1	Cytoplasmic face of the membrane
426	UvrABC system protein B	P0A8F8	1	Cytoplasmic
427	ATP-independent RNA helicase dbpA	P21693	1	Cytoplasmic
428	ABC transporter ATP-binding protein yjjK	P37797	1	
429	Putative HTH-type transcriptional regulator yafC	P30864	1	
430	Side tail fiber protein homolog from lambdoid prophage Rac	P76072	1	
431	Phenylalanyl-tRNA synthetase alpha chain	P08312	1	Cytoplasmic
432	Glycine betaine/L-proline transport system permease protein proW	P14176	1	Integral membrane protein, inner membrane
433	Probable ATP-dependent transporter sufC	P77499	1	
434	Protein sirB1	P20101	1	
435	Hypothetical protein yigB	P23306	1	
436	Membrane-associated protein uidC	Q47706	1	
437	Lipoate-protein ligase A	P32099	1	Cytoplasmic

438	Hypothetical GST-like protein yliJ	P75805	1	
439	Sensor protein zraS	P14377	1	
440	Cytochrome bd-II oxidase subunit I	P26459	1	
441	Outer membrane usher protein CS3-2	P15484	1	
442	Probable crotonobetaine/carnitine-CoA ligase	P31552	1	
443	ATP-dependent hsl protease ATP-binding subunit hslU	P0A6H5	1	Cytoplasmic
444	Glucosamine--fructose-6-phosphate aminotransferase	P17169	1	Cytoplasmic
445	DNA-directed RNA polymerase alpha chain	P0A7Z4	1	
446	Probable phosphoglycerate mutase gpmB	P0A7A2	1	
447	Serine hydroxymethyltransferase	P0A825	1	Cytoplasmic
448	Glutamyl-tRNA synthetase	P04805	1	Cytoplasmic
449	Glutamate 5-kinase	P0A7B5	1	Cytoplasmic
450	Succinyl-CoA synthetase beta chain	P0A836	1	
451	Tryptophan synthase beta chain	P0A879	1	
452	Adenylate cyclase	P00936	1	Cytoplasmic
453	Biosynthetic arginine decarboxylase	P21170	1	Periplasmic
454	Tyrosine recombinase xerC	P0A8P6	1	Cytoplasmic; associated with DNA
455	High affinity ribose transport protein rbsD	P04982	1	Inner membrane-associated
456	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	P0A7F9	1	Cytoplasmic
457	50S ribosomal protein L18	P0C018	1	
458	S-ribosylhomocysteine lyase	Q8FEP8	1	
459	UDP-N-acetylmuramate--L-alanine ligase	P17952	1	Cytoplasmic
460	ATP-dependent Clp protease ATP-binding subunit clpX	P0A6H1	1	
461	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	P62623	1	
462	Hypothetical protein yghW	P64574	1	
463	Glutamate-1-semialdehyde 2,1-aminomutase	P23893	1	Cytoplasmic
464	Trigger factor	P0A850	1	
465	Hypothetical UPF0250 protein ybeD	P0A8J4	1	
466	Holliday junction DNA helicase ruvB	P0A812	1	
467	Agmatinase	P60651	1	
468	N-methyl-L-tryptophan oxidase	P40874	1	
469	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	P62707	1	
470	Small heat shock protein ibpB	P0C058	1	
471	Hypothetical acetyltransferase ypeA	P76539	1	

472	Phosphopentomutase	P0A6K6	1	Cytoplasmic
473	Aspartate 1-decarboxylase	P0A790	1	
474	Cell division protein ftsB	P0A6S5	1	Type II membrane protein, inner membrane
475	Hypothetical oxidoreductase ybiC	P30178	1	Cytoplasmic
476	Hypothetical UPF0231 protein yacL	P0A8E5	1	
477	30S ribosomal protein S21	P68679	1	
478	Cysteine desulfurase	P0A6B7	1	
479	Regulator of ribonuclease activity A	P0A8R0	1	
480	Imidazole glycerol phosphate synthase subunit hisH	P60596	1	
481	6,7-dimethyl-8-ribityllumazine synthase	P61715	1	
482	Citrate synthase	P00891	1	
483	DNA recombination protein rmuC	P27850	1	
484	Lipoprotein releasing system transmembrane protein lolC	P75956	1	Integral membrane protein, inner membrane
485	Biopolymer transport exbB protein	P18783	1	Integral membrane protein, inner membrane
486	Cytochrome d ubiquinol oxidase subunit II	P11027	1	Integral membrane protein, inner membrane
487	Hypothetical UPF0053 protein yegH	P76389	1	Integral membrane protein
488	Hypothetical metabolite transport protein yhjE	P37643	1	Integral membrane protein, inner membrane
489	Protein glpG	P09391	1	
490	Putative permease perM	P77406	1	Integral membrane protein
491	Methyl-accepting chemotaxis protein II	P07017	1	Integral membrane protein, inner membrane
492	Type I restriction enzyme EcoKI R protein	P08956	1	
493	Hypothetical protein ycbC	P36565	1	
494	Small protein A	P23089	1	Attached to the outer membrane by a lipid anchor
495	Histidine biosynthesis bifunctional protein hisB	P06987	1	Cytoplasmic
496	DcrB protein	P37620	1	Periplasmic
497	Hypothetical protein yhfK	P45537	1	Integral membrane protein
498	Oligopeptide transport ATP-binding protein oppF	P77737	1	Inner membrane-associated
499	Hypothetical protein yddV	P77793	1	
500	2-oxoglutarate dehydrogenase E1 component	P07015	1	
501	Peptide transport system ATP-binding protein sapD	P36635	1	Inner membrane-associated
502	Lipoprotein releasing system transmembrane protein lolE	P75958	1	Integral membrane protein, inner membrane
503	Dipeptide transport system permease protein dppC	P37315	1	Integral membrane protein, inner membrane
504	Hypothetical protein yedI	P46125	1	Integral membrane protein
505	TolA protein	P19934	1	Type II membrane protein, inner membrane

506	Phosphatase ybhA	P21829	1	
507	Arginine transport ATP-binding protein artP	P30858	1	Inner membrane-associated
508	6-phosphogluconate dehydrogenase, decarboxylating	P37754	1	
509	Hypothetical ABC transporter ATP-binding protein yejF	P33916	1	
510	50S ribosomal protein L13	P02410	1	
511	Universal stress protein E	P03807	1	Cytoplasmic
512	Aspartate aminotransferase	P00509	1	Cytoplasmic
513	Putative sensor-like histidine kinase yfhK	P52101	1	Integral membrane protein, inner membrane
514	Hypothetical protein ybaL	P39830	1	Integral membrane protein
515	Lipopolysaccharide core biosynthesis protein rfaP	P25741	1	
516	Sensor protein rstB	P18392	1	Integral membrane protein, inner membrane
517	1,4-dihydroxy-2-naphthoate octaprenyltransferase	P32166	1	Integral membrane protein, inner membrane
518	Hypothetical transport protein yeeF	P33016	1	Integral membrane protein, inner membrane
519	Hypothetical UPF0105 protein yfcH	P77775	1	
520	Hypothetical protein yedE	P31064	1	Integral membrane protein
521	Hypothetical protein ymdC	P75919	1	
522	RNA polymerase sigma-E factor	P34086	1	
523	Mlc protein	P50456	1	
524	Hypothetical protein yrbK	P45397	1	
525	3-oxoacyl-[acyl-carrier-protein] reductase	P25716	1	
526	HTH-type transcriptional regulator cysB	P06613	1	Cytoplasmic
527	50S ribosomal protein L21	P02422	1	
528	TolR protein	P05829	1	Type II membrane protein, inner membrane
529	Hypothetical UPF0141 protein yjdB	P30845	1	Integral membrane protein
530	Putrescine transport ATP-binding protein potG	P31134	1	Inner membrane-associated
531	Maltose transport system permease protein malF	P02916	1	Integral membrane protein, inner membrane
532	Rtn protein	P76446	1	Membrane-associated
533	Hypothetical protein yaiW	P77562	1	
534	Hypothetical protein yjdA	P16694	1	
535	Pyridoxamine 5'-phosphate oxidase	P28225	1	
536	Alkyl hydroperoxide reductase subunit C	P26427	1	
537	Hypothetical protein yfhB	P30133	1	
538	Probable csgAB operon transcriptional regulatory protein	P52106	1	
539	Cell division protein ftsW	P16457	1	Integral membrane protein, inner membrane

540	Putative aliphatic sulfonates transport permease protein ssuC	P75851	1	Integral membrane protein, inner membrane
541	Hypothetical UPF0003 protein ynaI	P77253	1	Integral membrane protein
542	Hypothetical protein yddB	P31827	1	
543	H(+)/Cl(-) exchange transporter clcA (ClC-ec1)	P37019	1	Integral membrane protein, inner membrane
544	Hypothetical protein yegE	P38097	1	
545	Hypothetical protein yhhM	P37615	1	
546	Outer membrane usher protein papC	P07110	1	Integral membrane protein, outer membrane
547	Aminopeptidase N	P04825	1	Cytoplasmic, bound to the inner face of the cytoplasmic membrane
548	Aldehyde-dehydrogenase like protein yneI	P76149	1	
549	Carbohydrate diacid regulator	P37047	1	
550	50S ribosomal protein L14	P02411	1	
551	Hypothetical protein yccS	P75870	1	Integral membrane protein
552	Adenylosuccinate lyase	P25739	1	
553	Uracil permease	P33780	1	Integral membrane protein, inner membrane
554	Hypothetical protein yqjA	P42614	1	Integral membrane protein
555	Glutamate/aspartate transport ATP-binding protein gltL	P41076	1	Inner membrane-associated
556	Membrane-protein yhjW	P37661	1	Integral membrane protein
557	Biopolymer transport exbD protein	P18784	1	Type II membrane protein, inner membrane
558	RepA protein	P05833	1	
559	Flagellar fliL protein	P06973	1	Inner membrane-associated
560	ATP-dependent Clp protease ATP-binding subunit clpA	P15716	1	
561	Galactitol-1-phosphate 5-dehydrogenase	P37190	1	
562	Phosphatidylglycerophosphatase B	P18201	1	Integral membrane protein, Inner and outer membranes
563	Hypothetical protein yicN	P31439	1	
564	Cell division protein ftsQ	P06136	1	Type II membrane protein, inner membrane
565	NADH-quinone oxidoreductase chain E	P33601	1	
566	Hypothetical protein yffR	P76549	1	
567	Hypothetical protein ybjT	P75822	1	
568	Protein perC	P43475	1	
569	Polysialic acid biosynthesis protein P7	Q47400	1	
570	Outer membrane usher protein aggC	P46005	1	Integral membrane protein, outer membrane
571	Mrr restriction system protein	P24202	1	

572	Hypothetical oxidoreductase ydhF	P76187	1	
573	D-allose transport ATP-binding protein alsA	P32721	1	Inner membrane-associated
574	L-aspartate oxidase	P10902	1	Cytoplasmic
575	HTH-type transcriptional regulator betI	P17446	1	
576	Hypothetical UPF0118 protein ydiK	P77175	1	Integral membrane protein
577	TraD protein	P22708	1	Integral membrane protein, inner membrane
578	Glycolate oxidase subunit glcD	P52075	1	
579	Hypothetical protein yfeY	P76537	1	
580	D-serine/D-alanine/glycine transporter	P39312	1	Integral membrane protein, inner membrane
581	Hydrogen peroxide-inducible genes activator	P11721	1	
582	FolD bifunctional protein	P24186	1	
583	Transcriptional activator perA	P43459	1	
584	Putative HTH-type transcriptional regulator ycjW	P77615	1	
585	Macrolide-specific ABC-type efflux carrier	P75831	1	Integral membrane protein, inner membrane
586	Cystathionine gamma-synthase	P00935	1	Cytoplasmic
587	Ribonucleoside-diphosphate reductase 1 alpha subunit	P00452	1	
588	Hypothetical 11.1 kDa protein	P11907	1	
589	Galactoside transport system permease protein mglC	P23200	1	Integral membrane protein, inner membrane
590	ABC transporter ATP-binding protein yojI	P33941	1	Integral membrane protein, inner membrane
591	Fumarate reductase flavoprotein subunit	P00363	1	
592	HdeD protein	P26603	1	
593	Hypothetical outer membrane usher protein yqiG	P76655	1	Integral membrane protein, outer membrane
594	Cytosine permease	P25525	1	Integral membrane protein, inner membrane
595	Cytochrome c biogenesis ATP-binding export protein ccmA	P33931	1	Inner membrane-associated
596	Protoporphyrinogen oxidase	P27863	1	
597	Hypothetical protein yjbM	P32694	1	
598	Hypothetical protein yeaJ	P76237	1	
599	Hypothetical lipoprotein ygdI	P65292	1	Attached to the membrane by a lipid anchor
600	30S ribosomal protein S13	P0A7S9	1	
601	Hypothetical UPF0194 membrane protein ybhG	P75777	1	Membrane-associated Inner membrane
602	Thiol:disulfide interchange protein dsbD	P36655	1	Integral membrane protein, inner membrane
603	Glycyl-tRNA synthetase beta chain	P00961	1	Cytoplasmic
604	TolB protein	P0A855	1	Periplasmic
605	Transaldolase B	P0A870	1	Cytoplasmic

606	Polyphosphate kinase	P0A7B1	1	Inner membrane-associated
607	Glucosamine-6-phosphate deaminase	P0A759	1	
608	Nickel/cobalt efflux system rcnA	P76425	1	Integral membrane protein
609	Cation/acetate symporter actP	P32705	1	Integral membrane protein
610	Elongation factor Ts	P0A6P1	1	Cytoplasmic
611	Sec-independent protein translocase protein tatE	P0A843	1	Inner membrane-bound
612	Phosphoadenosine phosphosulfate reductase	P17854	1	Cytoplasmic
613	Succinate dehydrogenase cytochrome b556 subunit	P69054	1	Integral membrane protein, inner membrane
614	NAD-dependent deacetylase	P75960	1	Cytoplasmic
615	50S ribosomal protein L9	P0A7R1	1	
616	Aspartate carbamoyltransferase regulatory chain	P0A7F3	1	
617	Hypothetical UPF0259 protein yciC	P21365	1	Integral membrane protein
618	Glutathione-regulated potassium-efflux system protein kefC	P03819	1	Integral membrane protein, inner membrane
619	Cell division topological specificity factor	P0A734	1	
620	Cytochrome c-type protein torC	P33226	1	Type II membrane protein, inner membrane
621	Putative transport protein yidE	P60872	1	Integral membrane protein
622	Chromosome initiation inhibitor	P0A8S1	1	
623	50S ribosomal protein L11	P0A7J7	1	
624	Aspartate carbamoyltransferase catalytic chain	P0A786	1	
625	Hypothetical UPF0145 protein ybjQ	P0A8C1	1	
626	Probable glutamate/gamma-aminobutyrate antiporter	P63235	1	Integral membrane protein, inner membrane
627	DnaJ-like protein djlA	P31680	1	Type III membrane protein
628	Glutamyl-tRNA reductase	P0A6X1	1	
629	Ribosomal large subunit pseudouridine synthase B	P37765	1	
630	Cyclic di-GMP binding protein	P37652	1	
631	Alpha,alpha-trehalose-phosphate synthase	P31677	1	
632	Phosphoribosylaminoimidazole-succinocarboxamide synthase	P0A7D7	1	
633	Spermidine/putrescine import ATP-binding protein potA	P69874	1	Inner membrane-associated
634	Hypothetical UPF0114 protein yqhA	P67244	1	Integral membrane protein
635	Hypothetical protein ydaM	P77302	1	
636	Hypothetical protein yoaF	P64493	1	
637	Arginine repressor	P0A6D0	1	Cytoplasmic
638	Protein pnuC.	P31215	1	Integral membrane protein, inner membrane
639	Taurine import ATP-binding protein tauB	Q47538	1	Inner membrane-associated
640	Cardiolipin synthetase	P0A6H8	1	Membrane-bound

Chapter 8

Conclusions and Future Work

This thesis has focused on two aspects of current proteomics research: (1) development of a novel method for characterization of post-translational modifications (PTMs), and (2) development and applications of simple, global, economical, efficient and reliable quantitative proteomics strategies based on the integration of microbore LC-MALDI and stable isotope labeling.

In Chapter 2, the development of a novel approach for the characterization of PTMs of low-mass proteins observed in MALDI MS spectra is described. Using this approach, not only previously reported PTMs involving acetylation, methylation, oxidation and the removal of signal peptides, but also several novel PTMs, such as loss of N-terminal Met-Thr-Met (MTM) and hydroxylation of arginine, were identified. While the current method was developed and applied to the analysis of PTMs of proteins extracted from a relatively well-characterized microorganism, (*E. coli*), it is envisaged that this method should be applicable to other biological systems for the discovery of new PTMs.

Chapter 3 describes the development of an approach to global quantitative analysis of protein mixtures using differential stable isotopic labeling of enzyme-digested peptides combined with microbore LC-MALDI MS. In this work, microbore LC was combined with MALDI MS via a heated droplet interface. The compatibilities of two global peptide labeling methods (i.e., esterification of carboxylic groups and dimethylation of amine groups of peptides) with this LC-MALDI technique were evaluated. Using a QqTOF MS, MALDI spectra of the peptides in individual sample spots were obtained to determine the abundance ratio among pairs of differentially isotopically labeled peptides. MS/MS spectra were subsequently obtained from the peptide pairs to determine the sequences of selected peptides for protein identification. The peptide sequences determined from MS/MS database searches were confirmed by using the overlaid fragment ion spectra generated from a pair of differentially labeled peptides. Microbore LC provides higher sample loading, compared to capillary LC, and gives rise to accurate protein quantification due to the increased signal-to-noise ratio for a protein mixture with a concentration dynamic range of as high as 1×10^4. The effectiveness of this microbore LC-MALDI approach was demonstrated in the quantification and identification of peptides from a mixture of standard proteins, as well as an *E. coli* whole cell extract of known relative concentrations. It was shown that this approach provided a facile and economical means of comparing relative protein abundances from two proteome samples.

The purpose in developing a quantitative proteomics strategy was to quantify and identify biologically interesting proteins, thereby assisting biologists to a better understanding of their biological systems of interest. In Chapter 4, the deployment of a quantitative proteomics strategy based on the dimethyl labeling and microbore LC-MALDI QqTOF MS, described in Chapter 3, to quantify accurately and confidently identify differentially expressed proteins between an E-cadherin-deficient human carcinoma cell line (SCC9) and its transfectants expressing E-cadherin (SCC9-E) is described. A total of 5480 peptide pairs were examined and 320 of them showed relative intensity changes of greater than 2-fold. MS/MS analysis of these pairs led to the identification of 49 differentially expressed proteins between the parent SCC9 cells and SCC9-E transfectants. These proteins were determined to be involved in different pathways regulating cytoskeletal organization, cell adhesion, epithelial polarity and cell proliferation. The changes in protein expression were consistent with increased cell-cell and cell-matrix adhesion and decreased proliferation in SCC9-E cells, in line with E-cadherin tumor suppressor activity. Finally, the accuracy of the MS quantification and subcellular localization for 6 differentially expressed proteins were validated by immunoblotting and immunofluorescence assays.

Although a quantitative proteomics strategy based on differential dimethyl labeling using d(0)- or d(2)-formaldehyde has been developed (Chapter 3), and successfully applied to quantify and identify biologically interesting proteins between parent SCC9 cells and SCC9-E transfectants (Chapter 4), this differential dimethyl labeling method using d(0)- or d(2)-formaldehyde has the following shortcomings for bottom-up proteomics. First, the difference of 4 Da between peptide pairs with a mass of greater than 1900 Da leads to a significant overlap of the isotope envelopes for peptide pairs. Therefore, to generate accurate quantification results, additional peak deconvolution using an appropriate software algorithm is required, which complicates the data analysis process. Second, missed cleavages are observed very often when trypsin is used as the enzyme, which results in 1 to 3 lysine residues occurring frequently in the tryptic peptides. Thus, multiple labeling of tryptic peptides is often observed and it complicates the selection of peptide pairs for quantification and collision-induced dissociation (CID). This is particularly true for an LC-MALDI based method in which selection of proper peptide pairs is critical for abundance-ratio-dependent quantitative analysis. Therefore, the work described in Chapter 5 was devoted to developing a modified N-terminal dimethyl labeling strategy, in which the N-termini of tryptic peptides were differentially labeled with either d(0),^{12}C-formaldehyde or d(2),^{13}C-formaldehyde after lysine residues in peptides were blocked by guanidination. A nominal mass difference of 6 Da between the peptide pair allows negligible interference between the two isotopic clusters for quantification of peptides of up to 3000 Da. Since only the N-termini of tryptic peptides were differentially labeled and the a_1 ions were also enhanced in the MALDI MS/MS spectra, interpretation of the fragment ion spectra to obtain sequence information was greatly simplified. It was demonstrated that this technique of N-terminal dimethylation (2ME) after lysine guanidination (GA) or 2MEGA offered several desirable features, including simple experimental procedure, stable products, the use of inexpensive and commercially available reagents, and negligible isotope effect on reversed-phase separation. In addition to its applicability to quantitative analysis, LC-MALDI MS combined with this 2MEGA labeling technique was successfully used to identify polymorphic variants, phosphopeptides, and low abundance proteins in the whey fraction of bovine milk.

An attractive feature of the LC-MALDI-based quantitative proteomics strategy is its selective ability to quantify and identify differentially expressed proteins between two cell lines. The key step in using the integrated strategy of microbore LC-MALDI and 2MEGA labeling for abundance ratio dependent quantitative proteome analysis is to define the abundance ratio threshold. A very stringent threshold will lose important information to peak list generation and further MS/MS analysis. However, a very low threshold will result in wasting efforts on analyzing data with hardly any constructive meaning. In Chapter 6, the precision or reproducibility issue of the quantitative proteome analysis, i.e., how the technical variation would affect the quantitative results, has been addressed. All the measured CVs are well below the 15% that is normally employed as a threshold for bioanalysis, indicating that the developed quantitative proteome analysis strategy provides very good reproducibility. The biological variation originating from cell growth has been found to be less than the technical variation during the downstream protein analysis work, including protein assay, tryptic digestion, labeling, and microbore LC-MALDI mass spectrometric detection. The results have revealed that for future quantitative analysis of biologically significant proteins, more experimental attention should be paid to the analytical laboratory instead of the biological laboratory. Setting a relative abundance ratio of greater than 2-fold was demonstrated to be a very stringent threshold to quantify and identify differentially expressed proteins between two cultured cell lines. In practice, a 1.5-fold threshold can be employed to provide more detailed proteome profiles, avoiding the loss of some biologically interesting information when the integrated strategy is used for comparative proteome analysis. One of the greatest challenges associated with large-scale proteomics using tandem mass spectrometry (MS/MS) and automated database searching is how to reduce the number of false positive identifications without sacrificing the number of true positives found. In Chapter 7, a systematic investigation of the effect of 2MEGA labeling on the large-scale membrane proteome analysis by a shotgun proteomics strategy is detailed. By the large-scale comparison of MS/MS spectra from native peptides with those from the 2MEGA labeled peptides, the modified peptides were found to undergo facile fragmentation with signal-enhanced a_1 or a_1-related (a_1-17 and a_1-45) ions derived from all amino acids in the MS/MS spectra; ions that are usually difficult to detect in the MS/MS spectra of non-derivatized peptides. The 2MEGA labeling alleviated the biased detection of arginine-terminated peptides that is often observed in MALDI and ESI MS experiments. 2MEGA labeling was found not only to increase the number of peptides and proteins identified but also generate enhanced a_1 or a_1-related ions as a constraint to reduce the number of false positive identifications.

In conclusion, 2MEGA labeling combined with LC-MALDI MS and MS/MS or LC-ESI MS has been developed for quantitative proteome profiling. Quantitative proteomics has become one of the most important areas in the current proteomics field.

It is hoped that, with automation and further development of powerful separation techniques, many real world applications of the techniques developed in this thesis will follow. In Dr. Li's laboratory, other group members are developing multi-dimensional HPLC systems based on monolithic column technology and ultra-high pressure LC to improve the efficiency and speed of peptide and protein separations. These techniques will undoubtedly improve the overall performance of 2MEGA LC-MALDI or LC-ESI MS for quantitative profiling of a greater number of proteins from complex proteome samples. In addition, research on automation of the 2MEGA labeling process, which should benefit the analysis of a large number of samples, is being carried out. For real world applications, the integrated quantitative proteomics strategy developed in this thesis is being applied to several projects, including: (1) serum proteome analysis to search for potential biomarkers of breast cancer, (2) zebra fish liver proteome analysis to assess the functions and toxicity of environmentally significant chemicals, (3) membrane proteome profiling of cells grown under different conditions, (4) quantitative profiling of differentially expressed proteins between three breast cancer cell lines prepared under different conditions, and (5) heart tissue sample analysis to study the effects of drug application on heart function during surgery.